Modeling and Measuring Ultrasonic Nondestructive Evaluation Systems

This is a course organized by Lester W. Schmerr Jr., Professor Emeritus in Aerospace Engineering at Iowa State University. The content is based primarily on the book, **Ultrasonic Nondestructive Evaluation Systems – Models and Measurements**, authored by Prof. Schmerr and Prof. Sung-Jin Song, so that the reader is encouraged to purchase this book to enhance the value of the course.

The course contains annotated Power Point Slides that have been combined into this single pdf document that can be read on this website. The individual PowerPoints are also available for download.

Homework problems are also part of the course. The problem numbers are listed near the end of the sections where the appropriate topics are discussed. All of the course homework problems and solutions are available in a homework manual, which can either be read on this website or downloaded.

Matlab is used extensively in the course and homework to illustrate concepts and generate results. The Matlab course, homework, and book m-files are also available on this website for download.

The purpose of this course is to introduce the physics and mathematics behind ultrasonic flaw inspections so that one can understand the nature of the signals that are produced and how they are related to the flaws present. Linear systems theory, theory of elastic waves, and Fourier analysis serve as the foundations for much of the course. The course also covers briefly some aspects of statistical pattern recognition and probabilistic decision-making, and neural networks.

A Personal Preface (L.W. Schmerr)

As a graduate student in the Department of Mechanics at the Illinois Institute of Technology in the mid to late 60's I was trained in the area of elastic waves. After graduation, I went to Iowa State University in 1969 as an Assistant Professor in Engineering Mechanics. There, I started looking for a research area on which I could build a career. In the early 70's I participated in an NSF summer faculty program where I worked with the Nondestructive Testing group at General Dynamics, Fort Worth, Texas and where I was introduced to ultrasonics as a tool to inspect materials and to find flaws. At one point during the summer our group had a meeting with a General Dynamics corporate executive. After we had described our work, he asked us what we actually measured in an ultrasonic test. At that point in time I realized that there was simply not a good answer to his question and that finding one could be an important and interesting endeavor. Subsequently, modeling ultrasonic inspections became my main area of study which has endured to this day. In fact, this course is a snapshot of much of what I have learned over the years as to what we actually do measure in ultrasonic NDE.

As described in this course, we now do have reasonably good models of all the elements in an ultrasonic inspection. There were several key advances that serve as a foundation for these efforts. In 1979, Bert Auld from Stanford University developed a very general model of an inspection, based upon electrical and mechanical reciprocity principles, that could be used, in conjunction with elastic wave propagation and scattering models to predict ultrasonic signals received from flaws. However, while the Auld model is very general it does not relate those signals directly to any flaw properties being measured.

In 1983, Bruce Thompson and Tim Gray at the Center for NDE, Iowa State University, used the Auld model to develop a reduced model for "small" flaws (where the incident waves could be assumed to be a constant over the flaw geometry). This Thompson-Gray model, while less general than the Auld model, had the major advantage of explicitly delineating the flaw scattering response as a part of the entire ultrasonic measurement. Thompson and Gray also showed how this flaw response part (called the far field scattering amplitude) could be obtained by measuring what they called the "system efficiency factor" (in this course a very closely related factor is called the "system function") in a reference experiment, and then extracting the flaw response through deconvolution.

The Auld and Thompson-Gray models give us some general frameworks to model ultrasonic NDE inspections but there are many questions that remain. For example, how do the individual measurement system components (pulser/receiver, cabling, transducer(s)) affect the measured response? What are effective models to predict the waves generated by ultrasonic transducers? How is the far field flaw scattering amplitude obtained from the Thompson-Gray measurement model related to the actual geometry and material characteristics of the flaw? You will learn some answers to these and other important questions in this course.

The three books on modeling ultrasonic NDE inspections that I have written over the last 20 or so years can give you many of the modeling tools needed to predict flaw responses. However, the ability to extract specific flaw information (size, shape, properties, etc.) from those responses remains an area ripe for further investigation. Thus, at the end of this course I have included some lessons on statistical pattern recognition and neural networks. These areas can provide powerful tools for answering many of the very difficult questions that remain as to what we actually measure and how we can use those measurements to make rational engineering decisions from our ultrasonic NDE inspections.

I would like to close with a few words about my career. There are three accomplishments that I would like to highlight that I feel have contributed significantly to our understanding of ultrasonic tests. First, I reformulated Auld's model based on purely mechanical reciprocity principles. This is important since the elastic wave propagation and scattering elements in Auld's model are purely mechanical terms and by treating them as such one can see more clearly how they are imbedded in the entire measurement process through an acoustic/elastic transfer function. The details of this form of Auld's model are given in this course. I also showed how an ultrasonic transducer can be characterized as an electrical impedance and a sensitivity and how these factors can be obtained from a set of purely electrical measurements of the transducer in a pulse-echo setup. This greatly simplifies the previous procedures used in the acoustics literature which relied on a complex set of measurements involving multiple transducers. These details are also included in this course. In ultrasonic phased array inspections, one often uses those arrays to generate flaw images. A third accomplishment is not described in this course but is described in my book, Fundamentals of Ultrasonic Phased Arrays. There, I showed how three commonly used flaw imaging methods -the synthetic aperture focusing technique (SAFT), the total focusing method (TFM), and the Physical Optics Far-Field Scattering (POFFIS) method – are ad-hoc models that can be made more rational by inverting the Thompson-Gray measurement model to obtain a "Imaging Measurement Model". This model explicitly describes what these images generated by ultrasonic phased arrays represent in terms of the reflectivity and geometry of the flaws being imaged.

Although I have worked in ultrasonic NDE for many years, one of my most cited papers comes from my work on use of hypersingular integrals in the boundary element method (Krishnasamy, G., Schmerr, L.W., Rudolphi, T.J. and Rizzo, F.J., "Hypersingular boundary integral equations: some applications in acoustics and elastic wave scattering," Trans. ASME, Journ. of Appl. Mechanics, 57, 404-414, (1990).)

Although this paper is not on ultrasonic NDE per se, it is related since the integral equations governing the scattering of cracks are hypersingular. I will end by noting that is also somewhat ironic that the only paper that I have written that has received an award is not on ultrasonic NDE but on the use of neural networks in conjunction with eddy current inspections (American Society of Nondestructive Testing 1992 Achievement Award "in recognition of a manuscript that represents an outstanding contribution to the advancement of NDT" : Mann, J.M., Schmerr, L.W., and J.C. Moulder, "Inversion of eddy current data using neural networks," Materials Evaluation, 49, 34-39, 1991.)

The content of the course is mostly from my second book on ultrasonics but there are many results stated without proof in this course whose details can be found in my other two books. The references for all three books are:

Schmerr, L.W. and S.J. Song, *Ultrasonic Nondestructive Evaluation Systems – Models and Measurements*, Springer, 2007.
Schmerr, L.W., *Fundamentals of Ultrasonic Nondestructive Evaluation – A Modeling Approach*, 2nd Ed. Springer, 2016.
Schmerr, L. W., *Fundamentals of Ultrasonic Phased Arrays*, Springer, 2015.

TABLE OF CONTENTS

The topics of the course are listed below as well as the Chapters or Appendices of the book, **Ultrasonic Nondestructive Evaluation Systems – Models and Measurements** that correspond to the topics covered.

Introduction to NDE Ultrasonic Systems

This set of slides is a general introduction to the basic elements of an ultrasonic nondestructive evaluation (NDE) system designed to examine materials or structures for flaws.

Learning Objectives

familiarity with:
- ultrasonic system components
- ultrasonic NDE terminology

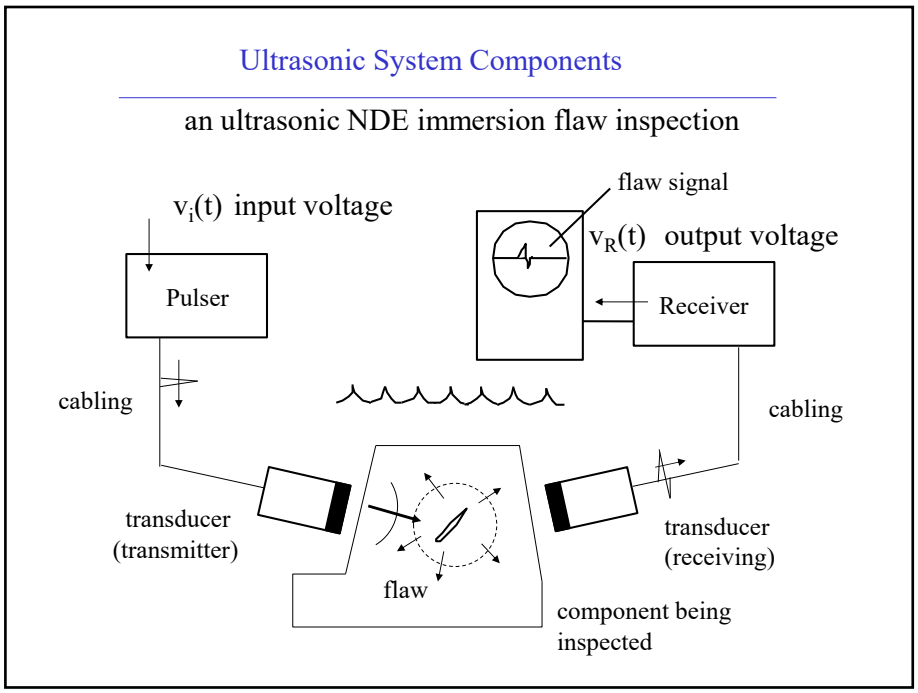
basic information on transducers
- construction
- types

common ultrasonic displays

We will examine the components that make up an ultrasonic NDE system and introduce you to some of the terminology that is commonly used in the ultrasonic NDE field.

Ultrasonic transducers are an important part of any system as they are used both to generate the ultrasound and to receive it. We will discuss some of the different types of transducers commonly used and how they are constructed.

There are various ways in which the signals received in an ultrasonic test are presented to the user so we will also describe some of the commonly used displays.



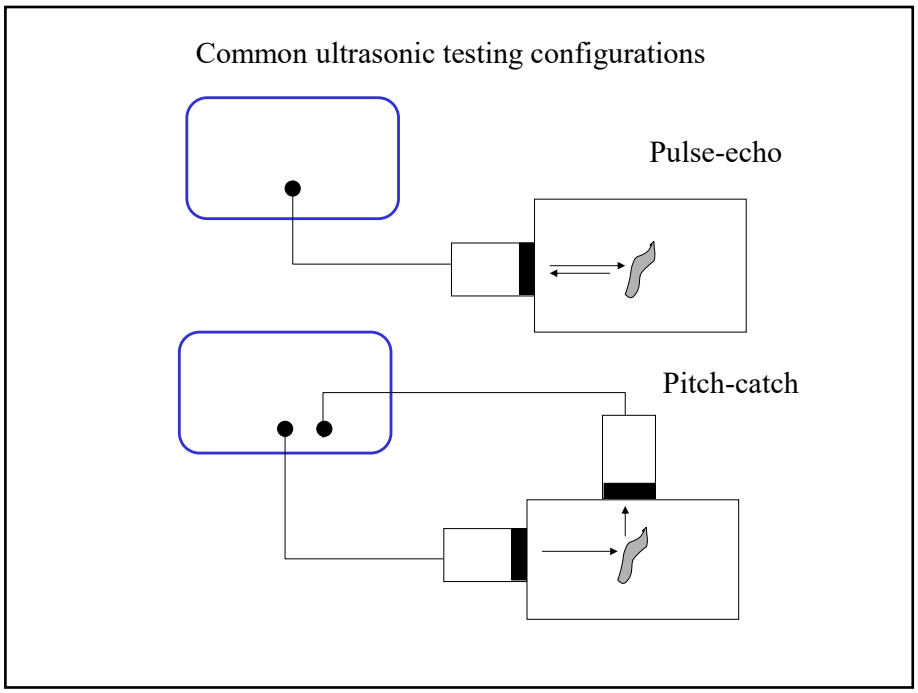
Here is a diagram of a basic ultrasonic NDE immersion system that uses two transducers placed in a water bath (one acts as a sound generator and one acts as a receiver) to examine a flawed component that is also in the water bath.

The pulser part of the system generates very short, repetitive voltage pulses (only one is shown) that travel through a cable and excite the sending transducer which converts those electrical pulses into pulses of sound that travel through the water and pass into the component.

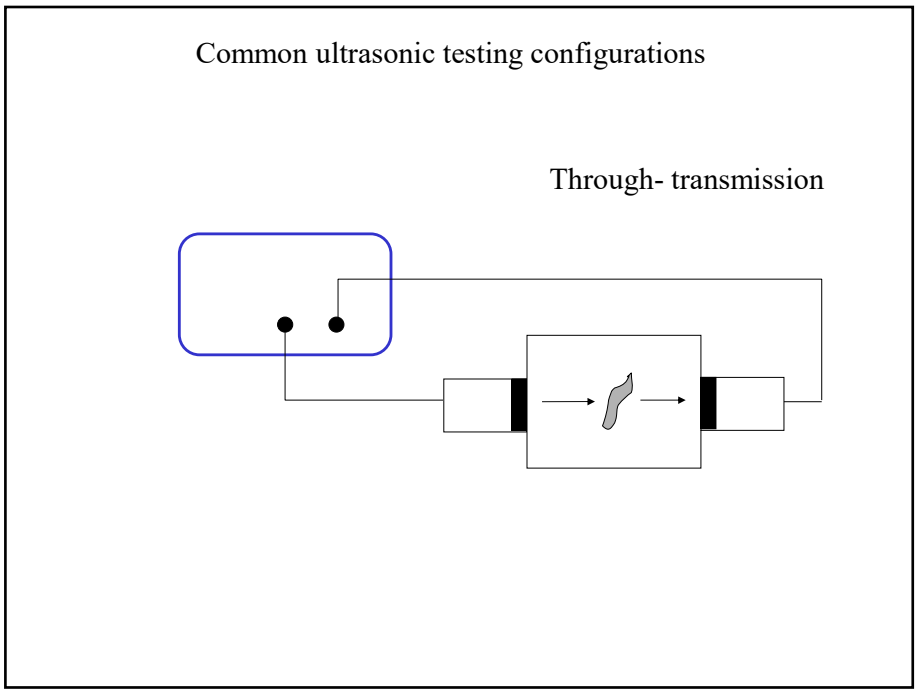
These incident waves in the component can interact with any flaws present. The flaws, in turn, generate additional scattered ultrasound waves which in this setup are picked up by a separate transducer and converted back to voltage pulses that travel over a cable to a receiver where they are amplified and displayed as a voltage versus time signal on an oscilloscope.

Pulser/Receiver

This is a picture of a very basic pulser/receiver used in a lab setting. We will examine the pulser/receiver in more detail later so here we will just note that the left hand side of the instrument contains the controls for the pulser which include the pulse repetition frequency and energy and damping settings while the receiver side of the instrument contains a gain setting and filtering options. The back of the instrument (not shown) contain connections for feeding the received signals to an oscilloscope display.

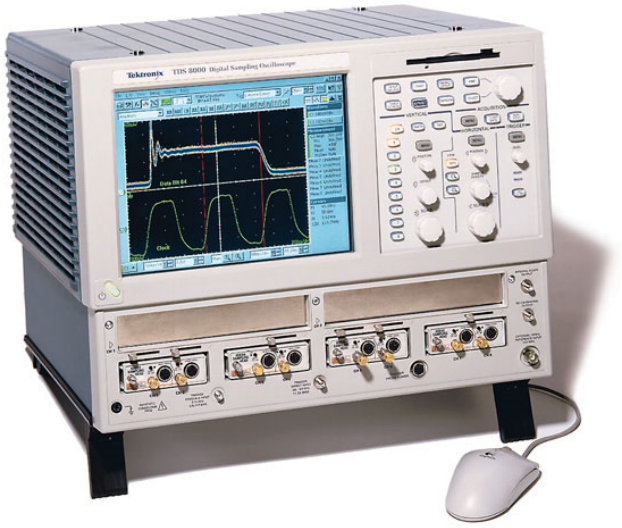


The pulser/receiver can either be used in a **pulse-echo** setting where a single transducer acts as both the transmitter and receiving transducer or in a **pitch-catch** setting where there are separate sending and receiving transducers.

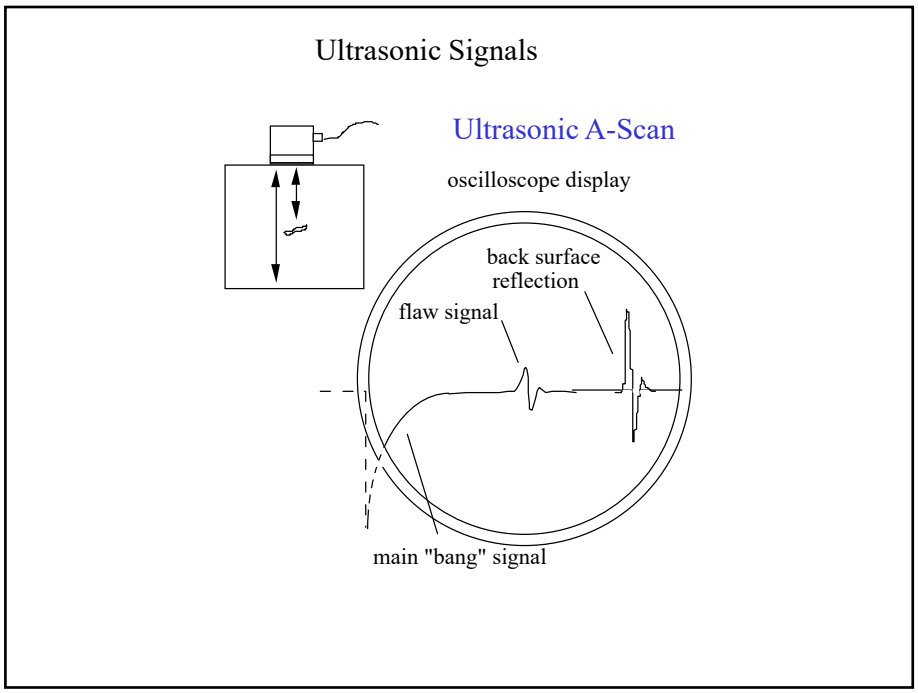


It is also possible to use two transducers that are directly opposite to each other in a **through-transmission** setup. In this case note that even with a flaw absent, the receiving transducer will still be receiving a signal so that the flaw in this case acts as a modifier of this signal that passes through the component. In the pitch-catch case, in contrast, if the flaw is absent no signal will be received (at least from interactions with the flaw).

Digital Oscilloscope



This is a picture of a typical digital oscilloscope which is used to digitally sample the received voltage versus time signals and display them on the oscilloscope screen. Since the signals are stored digitally, they can also be easily transferred to a computer for further processing.



A voltage versus time trace on an oscilloscope display is also called an **A-scan**. It is the most commonly used type of display in ultrasonic tests.

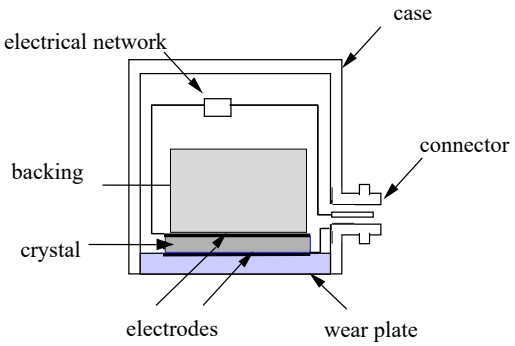
Here we see a transducer in contact with the surface of a part receiving signals from a flaw in the component as well as a reflected signal from the back surface in a pulse-echo setup. On the oscilloscope screen we see the received signals as well as a large negative pulse at the start of the display called the **main bang signal**. This is a portion of the driving pulse that leaks through the pulser-receiver circuits to receiving side. This large signal can hide any flaw signals that may be present that are very close to the transducer so that there is an early part of the display, called a "**dead zone**" , where evaluation of the received signals may not be possible.

Portable Flaw Detectors

Ultrasonic NDE tests are commonly done in the field, where a portable instrument is used. This instrument integrates the pulser/receiver and display into a single unit and provides the capability to perform various types of evaluations via the instrument settings.

Ultrasonic System Components

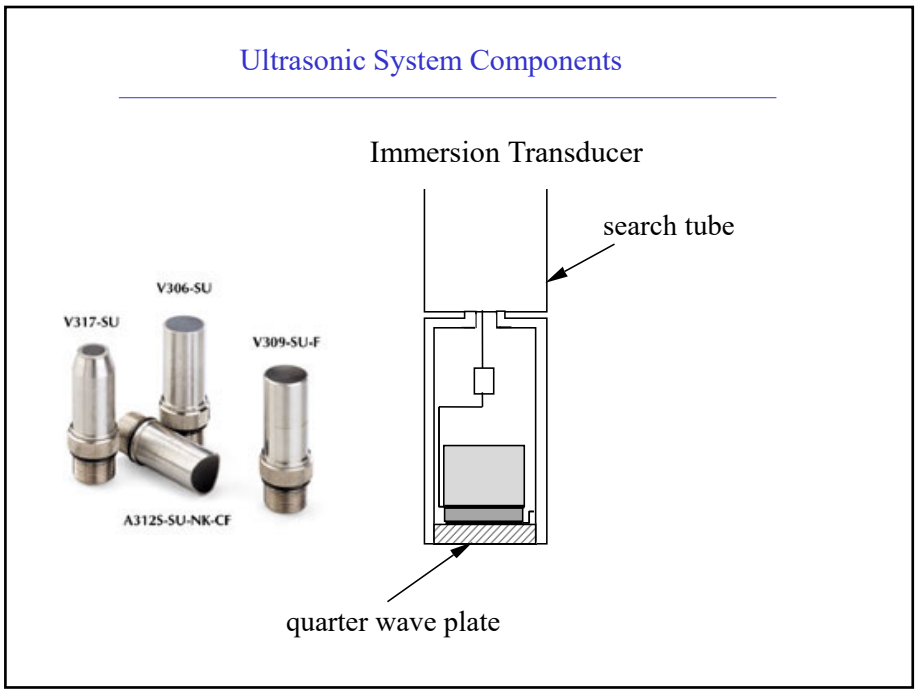
Contact Transducer



Here we see some typical ultrasonic transducers that are used in contact testing where the transducers are placed on the surface of a part and where a thin layer of water, oil, glycerin, or a commercially available couplant (not shown) is used to efficient couple the transducer to the part.

The transducer consists of a piezoelectric crystal which is plated on both faces. Acting as a transmitter, this crystal converts the electrical pulses acting on its faces into mechanical motion of the crystal, which then propagate through a wear plate and the couplant into the part as a traveling wave. The wear plate serves to protect the crystal from damage as the transducer is moved on the surface. A backing made of a highly attenuating material (such as epoxy loaded with small tungsten particles) dampens any waves traveling up into the transducer, thus preventing significant “ringing” of the crystal and generates a short pulse of ultrasound. There also may be some electrical components within the transducer casing that are used to adjust its electrical characteristics and “tune” the transducer output.

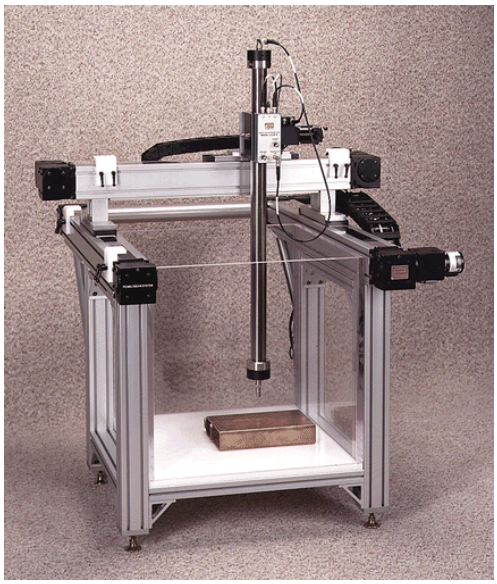
The same transducer can be used as a receiver which converts the motion of waves received into electrical signals.



Here we see some transducers that are used in an immersion setup.

The basic construction of an immersion transducer is very similar to a contact transducer except the transducer usually has a UHF type of connector that is screwed into a search tube in a scanning apparatus. Also, the wear plate is replaced by a so-called **quarter wave plate** which is specifically designed to efficiently couple the crystal to a water bath. Unlike the contact transducer, where it is difficult to maintain a constant coupling to the part as the transducer is moved around, in an immersion setup the water provides a constant coupling medium regardless of the motion of the transducer. Thus, immersion tests are more reproducible than contact tests.

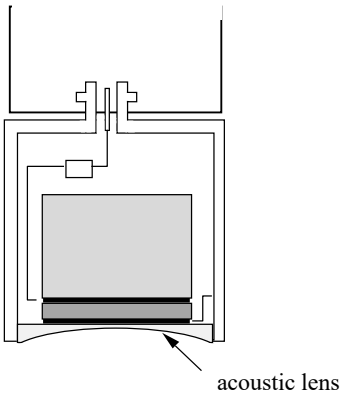
Immersion scanning system



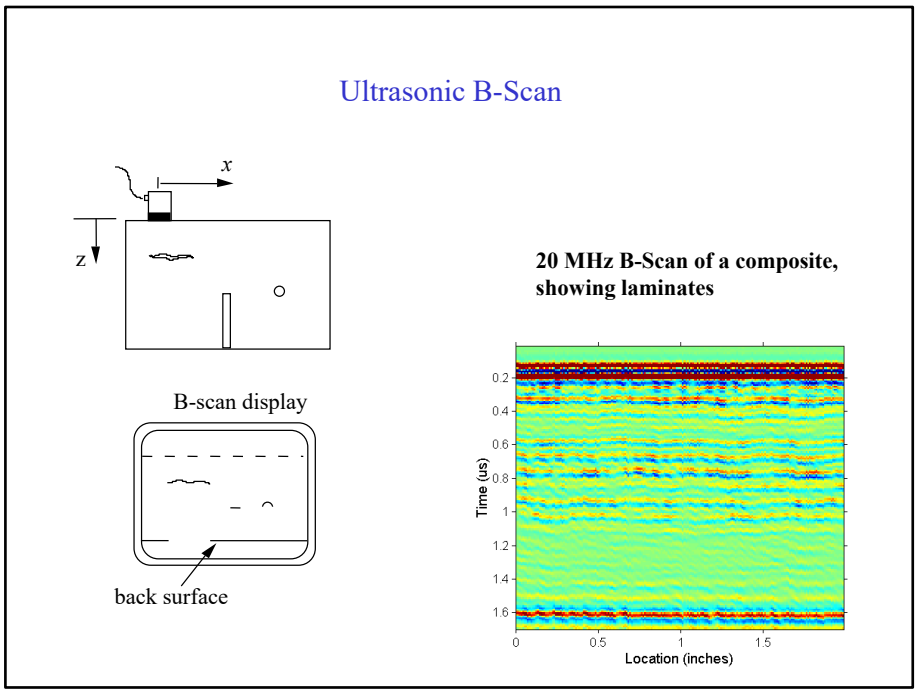
Here is a basic lab-top immersion setup where a transducer is connected to a scanning apparatus that allows the motion of the transducer to be controlled in a water tank. Much more sophisticated scanners can control both the angles and motions of the transducer.

Ultrasonic System Components

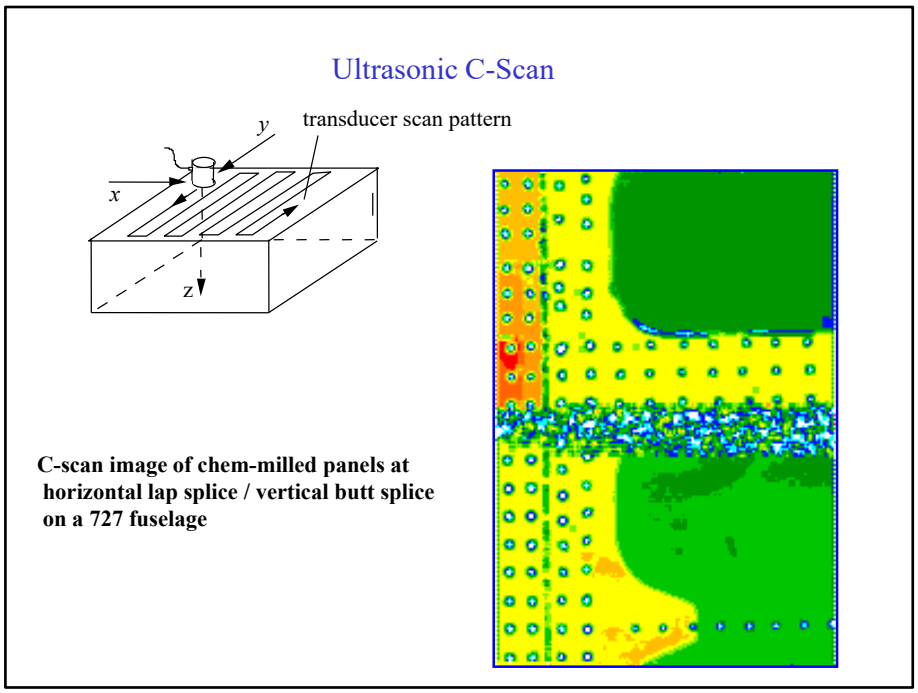
Focused Transducer



A planar (unfocused) immersion transducer generates a broad beam of ultrasound. However, if an acoustic lens is placed in front of the crystal, the sound can be focused into a part at a specific depth.



In addition to an A-scan, a **B-scan** is a commonly used type of display. Here, the x-location of the transducer is recorded as a transducer is moved along a line of the surface and the time of a received signal from a reflector is converted into the corresponding depth, z. A plot of x versus z on a display is called a B-scan. The picture on the right, for example, is a high frequency B-scan of a composite laminate where we can see the laminates clearly.



In a **C-scan** a transducer is scanned in two dimensions on a surface and the (x, y) location is recorded as well as the response from any subsurface reflectors, which is usually color coded as a function of the amplitude of the response. An example C-scan is shown of the C-scan of a lap splice for a Boeing 727 fuselage.

Ultrasonic System Components

References

Schmerr, L.W. and S.J. Song, **Ultrasonic Nondestructive Evaluation Systems – Models and Measurements**, Springer, 2007.

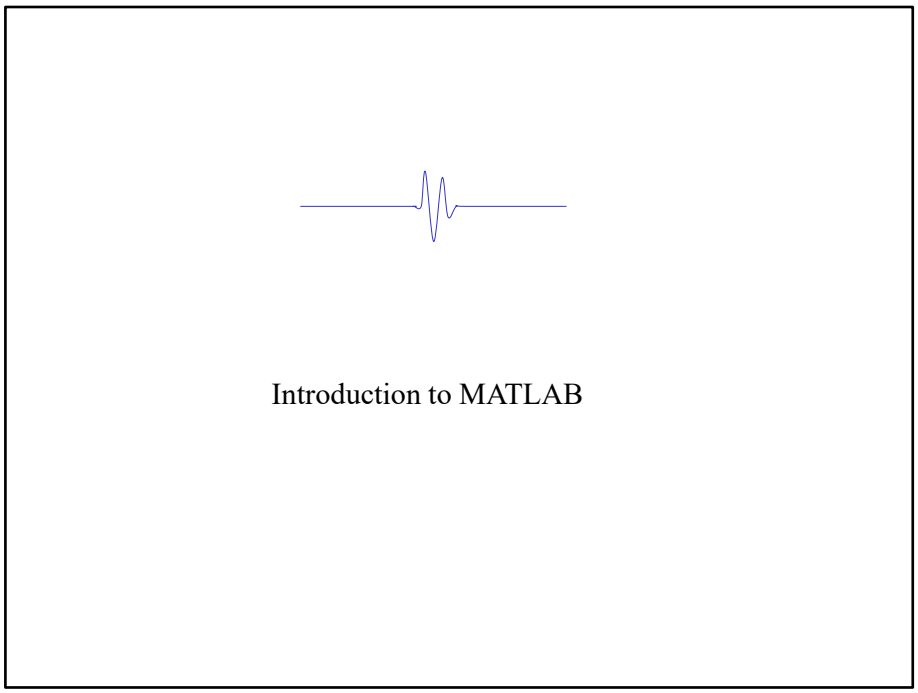
Schmerr, L.W., **Fundamentals of Ultrasonic Nondestructive Evaluation – A Modeling Approach, 2nd Ed.**, Springer, 2016.

Schmerr, L.W., **Fundamentals of Ultrasonic Phased Arrays**, Springer, 2015.

Krautkramer, J. and H. Krautkramer**, Ultrasonic Testing of Materials**, Springer Verlag, 1990.

Blitz, J. and G. Simpson, **Ultrasonic Methods of Non-destructive Testing**, Chapman & Hall, London,UK, 1996.

Here are the three ultrasonic NDE books published by the author (L. Schmerr) of these notes, as well as several other commonly used general references.



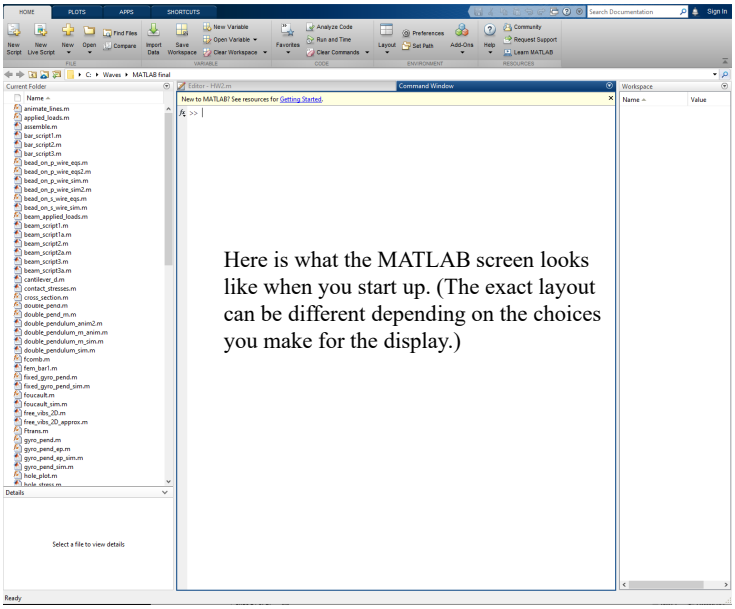
This set of slides gives a brief overview of MATLAB which is used often in these presentations to evaluate the models discussed. If you are very familiar with MATLAB you can skip this section. There are some homework problems at the end.

Learning Objectives

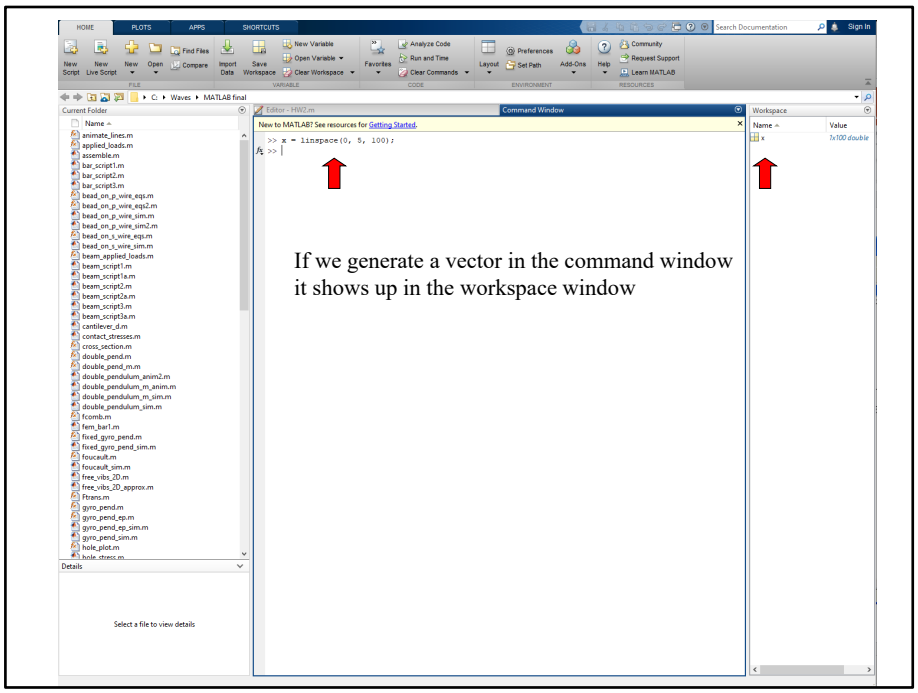
familiarity with:

- MATLAB operations
- Simple Plotting
- MATLAB functions, scripts
- Complex numbers
- Matrices, vectors

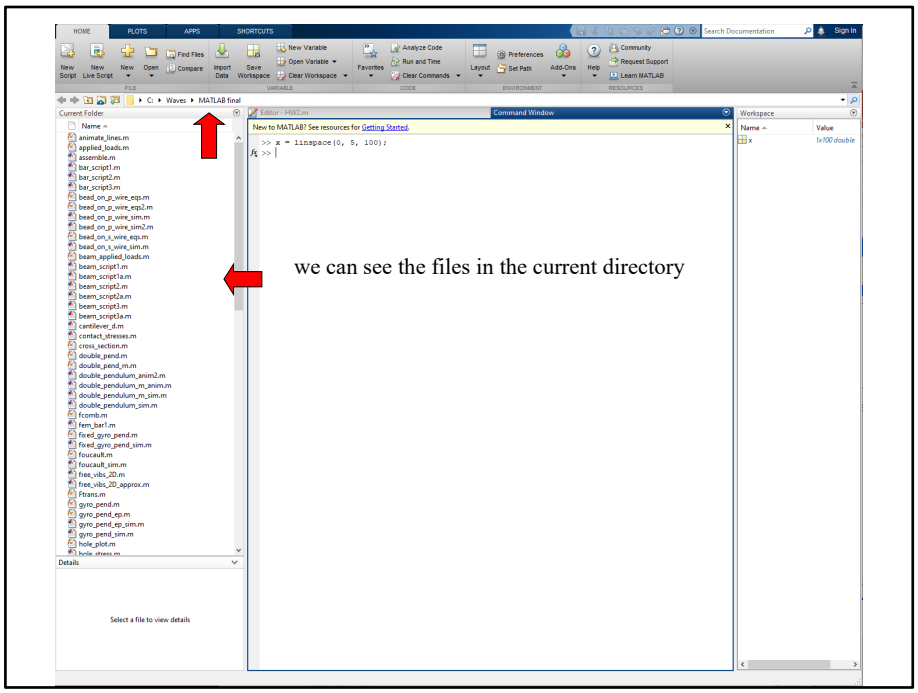
We will examine basic MATLAB operations, including simple plots. We will also discuss generating MATLAB functions and scripts. Inherently we will be working with vectors and matrices, which MATLAB is specifically designed to address, and we will be working with complex numbers.



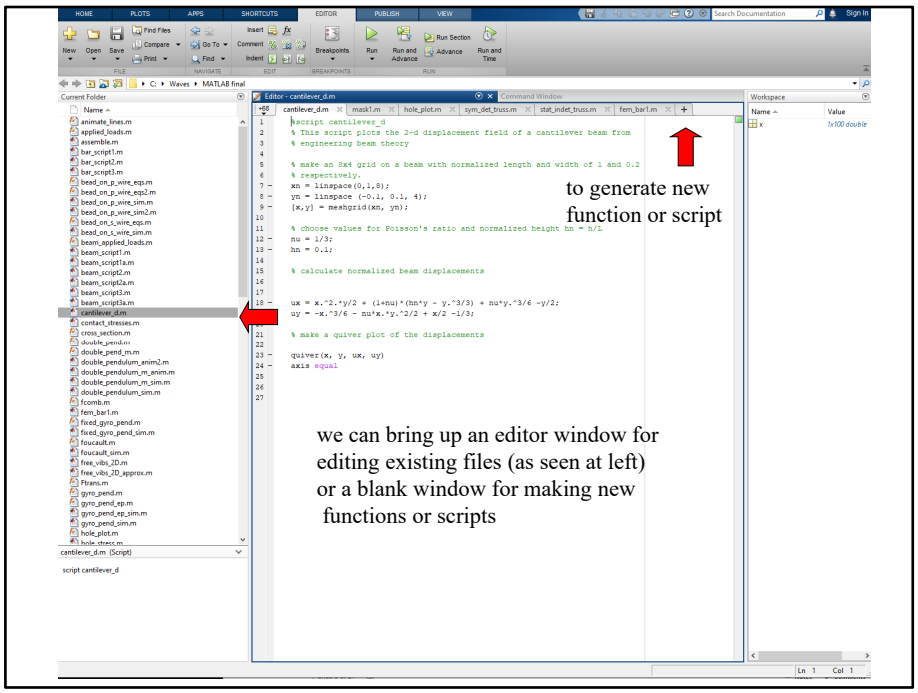
Here is a view of what my MATLAB screen looks like. The layout can be adjusted to show or hide various elements so your screen might not look identical to this one.



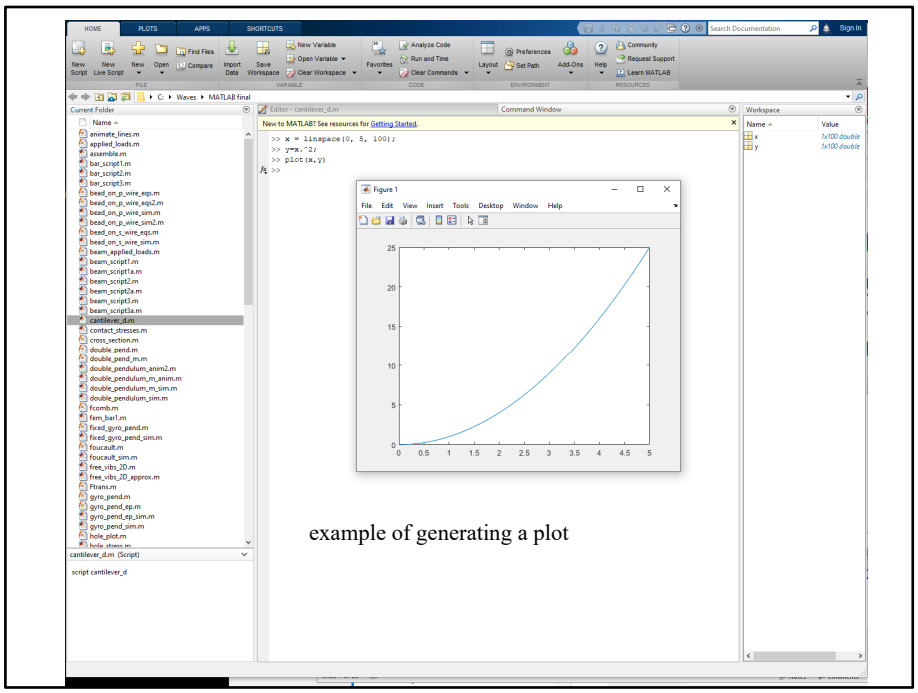
We can enter commands directly in the command window. Here we are generating a vector of 100 values ranging from zero to five. The resulting vector shows up in the workspace.



The contents of the current folder is shown in the left window.



If we select one of the m-files in the current folder, an editor window will replace the command window and show the contents of the selected file, which is in this case the script cantilever_d. If you click the plus sign at the upper right of the editor window it will bring up an empty editor window that you can use to generate a new script or function.



example of generating a plot

If you are in the editor and click the command window bar at the top, you will return to the command window. We can do simple plotting directly in the command window. Here we show the generation of the function $y = x^2$ in the command window, which is then plotted.

MATLAB

Built in constants and variables

ans	most recent answer
eps	small constant ~ 10^{-16}
i or j	imaginary unit
inf	infinity
pi	3.14159 ...

Standard mathematical operations

±	addition, subtraction
*	multiplication
/	division
^	exponentiation e.g. y^n = y^n

Some common functions

sin(x)	sine
cos(x)	cosine
exp(x)	exponential
sqrt(x)	square root
log(x)	natural log
log10(x)	log to base 10
abs(x)	absolute value, magnitude of complex quantity
angle(x)	phase angle
real(x)	real part of
imag(x)	imaginary part of

There are a number of built in constants and common functions. Here are a few. Also shown are some of the symbols used for standard operations.

Generation of (row) vectors

```
>>          x = 0:0.1:0.5
x =
            0    0.1000    0.2000    0.3000    0.4000    0.5000

>>          y =linspace(0, 0.5, 4)
y =
      0    0.1667    0.3333    0.5000

>>          z = [ 1  2  3  4  5]
z =
   1    2    3    4    5
```

Suppression of echoing of output (put semi-colon at end of line)

```
>>          z = [ 1  2  3  4  5] ;
>>
>>          z
z =
  1    2    3    4    5
```

Comment line

```
>>  %  This is a comment
>>
```

There are several ways to generate a vector of numbers. We can do this with the colon notation seen where start: sep :end generates numbers from start to end with a separation between values of sep. The function linspace(start, end, num) generates num equally spaced values going from start to end.

If we do not place a semi-colon at the end of the line, the result of the operation contained in the line will be echoed in the command widow as seen in a number of the above examples. A semi-colon at the end of the line suppresses that echoing.

To add comments to a line, which is very helpful for explaining your calculations, they must be preceded by the percent sign %

Functions can have vector (matrix) arguments

```
>>          x = [ 1  2  3  4  5];
>>          y  = exp(x)
y =
   2.7183    7.3891   20.0855   54.5982  148.4132
```

Element by element operations on vector-valued functions

±	addition, subtraction
.*	multiplication
./	division
.^	exponentiation

```
>>          x = [ 1  2  3  4]
x =
    1    2    3    4
>>          y = x + 2
y =
    3    4    5    6
>>          z = x .^2
z =
    1    4    9   16
>>          f = x .*x .^2
f =
    1    8   27   64
```

One of the nice things about MATLAB is that functions can take vectors or matrices in their arguments and produce a vector or matrix as their output. There are special symbols that allow us to do element by element operations on vector or matrix-valued functions. Some examples are shown.

MATLAB does complex arithmetic with scalars and vectors

```
>>              z = 3 + 4*i
z =
   3.0000+ 4.0000i
>>              y =i*z
y =
  -4.0000+ 3.0000i
>>              x =[ 1 + 3*i   2*i]
x =
   1.0000+ 3.0000i       0+ 2.0000i
>>y = [  i   i]
y =
       0+ 1.0000i       0+ 1.0000i
>>              z = x .*y
z =
  -3.0000+ 1.0000i    -2.0000
```

Common functions take complex arguments (scalars or vectors)

```
>>              exp(i*pi)
ans =
   -1
>>              x = [ 0   pi/2     pi] ;
>>              exp(i*x)
ans =
   1.0000     0+ 1.0000i         -1.0000
```

MATLAB recognizes i as the square root of minus one. One also can use 1i instead, which is helpful if we want to use i for other purposes. We can do complex arithmetic in the much same way we do real arithmetic and many common functions can take complex arguments.

Magnitude and phase of a complex number

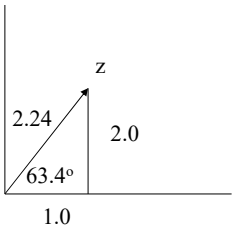
$$z = a + ib$$
$$= Ae^{i\phi}$$

$$A = |z| = \sqrt{a^2 + b^2}$$
$$\phi = \tan^{-1}(b/a)$$

In MATLAB:

```
>> z =1 + 2*i;
>> abs(z)
ans = 2.2361

>> angle(z)*180/pi
ans = 63.4349
```

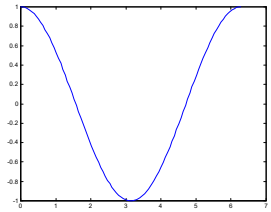
convert from radians to degrees



Complex numbers can be written as the sum of a real part and an imaginary part or they can be written as a magnitude times a complex exponential phase term. In MATLAB we can extract the magnitude of a complex number through the abs function and the phase through the angle function (which returns radians)

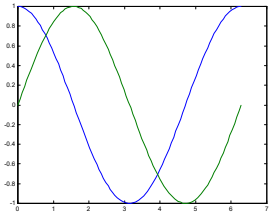
Simple plotting

```
>>      x = linspace( 0, 2*pi, 100);
>>      y = cos(x);
>>      plot(x, y)
```



Multiple plots on same graph

```
>>      x = linspace( 0, 2*pi, 100);
>>      y1 = cos(x);
>>      y2 = sin(x);
>>      plot(x, y1, x, y2)
```



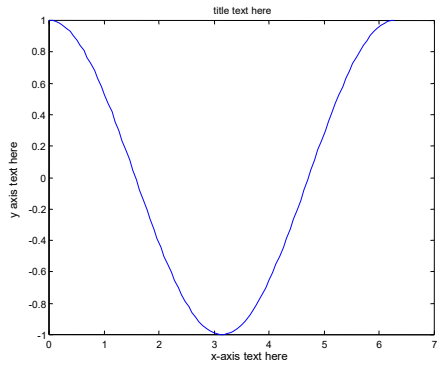
or

```
>>      x = linspace(0, 2*pi, 100);
>>      y1= cos(x);
>>      plot(x, y1)
>>      hold on
>>      y2 = sin(x);
>>      plot(x, y2)
>>      hold off
```

Here are some examples of doing simple plotting and putting multiple plots on the same graph.

Adding a x-axis label, a y-axis label, and a title to a plot

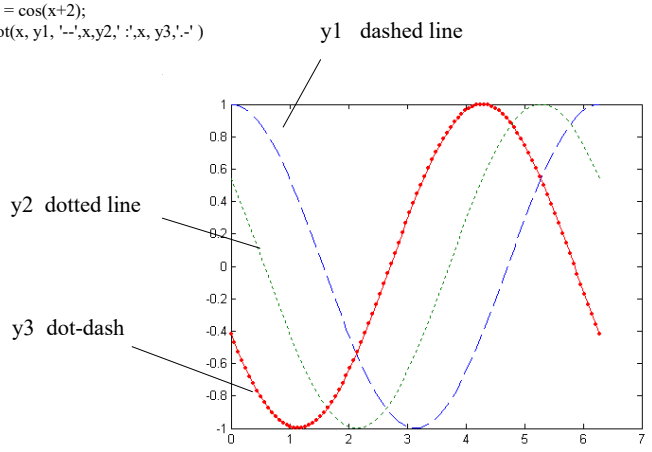
```
>>  x = linspace( 0, 2*pi, 100);
>>  y = cos(x);
>>  plot(x, y)
>>  xlabel(' x-axis text here')
>>  ylabel(' y axis text here')
>>  title('title text here')
```



We can also add x-axis and y-axis labels and give a plot a title. The arguments here are all strings which have the start and ending symbols ‘

Plotting with different line styles

```
>>  x = linspace( 0, 2*pi, 100);
>>  y1 = cos(x);
>>  y2 = cos(x+1);
>>  y3 = cos(x+2);
>>  plot(x, y1, '--',x,y2,' :',x, y3,'.-' )
```



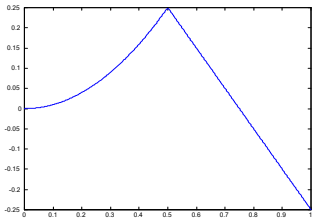
Here we see using strings for different plotting symbols to generate the plots shown.

Logical (0-1) vectors

```
>>          x = [ 1 2 3 4 5];
>>          x >3
ans =
    0    0    0    1    1
>>          x <= 4
ans =
    1    1    1    1    0
```

Use of logical vectors for defining piecewise function

```
>>          x = linspace(0, 1, 500);
>>          y =(x .^2).*(x  <  0.5) + (0.75 - x ).*(x  >=  0.5);
>>          plot(x, y)
```



A MATLAB statement such as x>3 for a vector x is treated as a logical statement and returns values of 0 in the vector when the statement is not true and 1 values when it is true. Logical vectors make it very easy to generate a function whose behavior changes depending on such a logical statement. Look at the example shown carefully to understand how the function plotted is generated from the ordinary functions x^2 and 0.75 –x with the use of logical vectors.

Defining parts of vectors

```
>>          x = [ 5 6 3 8 7 9];
>>          x(1:3)
ans =
    5    6    3
>>          x(2:4)
ans =
    6    3    8
>>          x(3:end)
ans =
3    8    7    9
```

Other vector operations

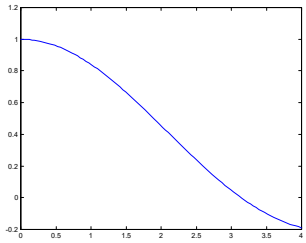
Vector magnitude, length

```
>>          x = [3 4];
>>          norm(x)
ans =
    5

>>          length(x)
ans =
    2
```

We can define various parts of vectors as shown in the above examples. We also determine the magnitude of a vector and the length of the vector (how many elements it contains)

Use of eps to avoid division by zero

```
>>          x=linspace(0, 4, 100);
>>          y=sin(x)./x;

Warning: Divide by zero.

>>          x = x + eps*(x = = 0);
>>          y=sin(x)./x;
>>          plot(x,y)
```



Use of inf to evaluate expression when a variable goes to infinity

```
function  y = infinitytest(x)
x = x + eps*( x = = 0);
y = 3/(1 + 4/x);
```

```
>>          infinitytest(1)
ans =
   0.6000
>>          infinitytest(0)
ans =
  1.6653e-16
>>          infinitytest(inf)
ans =
    3
```

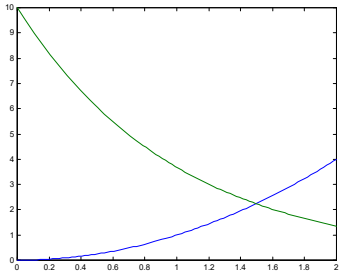
functions like sin(x)/x are formally 0/0 which is undefined at x = 0. However, since sin(x) goes like x when x is small, the limit of sin(x)/x at x =0 is 1. To get this proper limit in MATLAB we can add a small constant, eps, to the zero value to get the proper limit. Similarly, when we want the limit of an expression when an argument is very large (i.e. going to infinity), we can use the inf symbol in an expression to obtain that limit.

MATLAB Functions

Functions are defined in the editor and saved as m-files. The name of the m-file is the name of the function with a .m extension, e.g. myfunction.m etc. As shown in the example below, functions can have multiple outputs (as well as multiple inputs).

```
function  [y , z] = test(x)
y = x.^2;
z = 10*exp(-x);
```

```
>>    x = linspace(0, 2, 100);
>>    [s, t] = test(x);
>>    plot(x, s, x, t)
```



All the variables appearing in a function are local to that function, i.e. they do not change any similarly named variables in the MATLAB workspace.

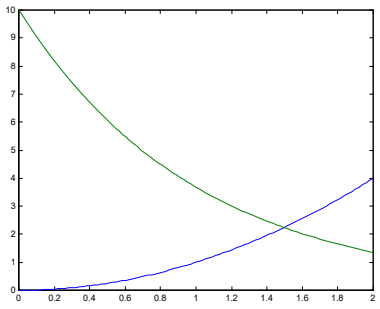
We create functions in the MATLAB editor window and then save them as m-files. We then can use them in the command widow, as shown, or in other functions or scripts. An important property of a function is that all the variables that are defined in the function are entirely separate from any similar variables in the workspace, i.e. they are local variables to that function.

MATLAB Scripts

A MATLAB script is a sequence of ordinary MATLAB statements, defined in the editor and then saved as a file.The file has the name of the script followed by a .m extension, e.g. myscript.m.
Typing the script name at the MATLAB prompt then causes the script to be executed. Variables in the script change any values of variables of the same name that exist in the current MATLAB session.

```
% testscript
x =linspace(0,2,100);
y = x.^2;
z = 10*exp(-x);
plot(x,y,x,z)
```

```
>>  testscript
```



A script can also be defined in the editor as simply a collection of MATLAB statements as they might be entered manually in the command window by the user. Variables defined in the script exist in the MATLAB workspace so they can use or overwrite existing workspace variables.

MATLAB Matrices

MATLAB was designed to perform operations with matrices very effectively. Entering matrices manually is as easy as vectors:

```
>> matrix  =  [  4 0 3; 0 3 5; 3 5 7  ]

matrix =
             4    0    3
             0    3    5
             3    5    7
```

Accessing individual components

```
>> matrix(2,3)
ans =
    5
```

Accessing rows or columns

```
>> matrix(: , 2)
ans =
             0
             3
             5
>> matrix(3, :)
ans =
        3    5    7
```

MATLAB can easily process matrices as well as vectors. Shown are a few of the common ways we can access the matrix contents.

Some matrix functions

size(M) returns number of rows, nr, and number of columns, nc, as a vector [nr , nc]
trace(M) trace of M (sum of diagonal terms)
det(M) determinant of M
M.' transpose of M (interchange rows and columns) If M is real then we can use M' instead. However, if M is complex then M' will interchange rows and columns and also perform a complex conjugation.

```
>>          size(matrix)
ans =
    3    3
>>          trace(matrix)
ans =
   14
>>          det(matrix)
ans =
  -43
>>          matrix'

ans =                                   (no change since matrix is symmetric)

    4    0    3
    0    3    5
    3    5    7
```

Here are a few of the functions available in MATLAB that operate on matrices.

Multiplying matrices and vectors

```
>>          v = [ 1 2 3];
>>          v*matrix
ans =

   13    21    34

>>          vt = v'
vt =

    1
    2
    3
>>          matrix*vt
ans =

   13
   21
   34
>>          v*matrix*vt

ans =

  157
```

transpose of vector v (turn row vector to column vector or vice-versa)

We can multiply matrices and vectors together. Here are some examples using the vector v and the previously defined matrix.

Special matrices

zeros(m,n)	matrix of all zeros with m rows and n columns
ones(m,n)	matrix of all ones with m rows and n columns
eye(m,n)	identity matrix with m rows and n columns

```
>>            zeros(3,3)
ans =
    0    0    0
    0    0    0
    0    0    0
>>            ones(3,3)
ans =
    1    1    1
    1    1    1
    1    1    1
>>            eye(3,3)
ans =
    1    0    0
    0    1    0
    0    0    1
```

Here are some special matrices than can be used as building blocks for various other matrices.

Solving a system of linear equations

$$1x_1 + 3x_2 + 5x_3 = 3$$
$$2x_1 + 1x_2 + 1x_3 = 2$$
$$4x_1 + 3x_2 + 6x_3 = 1$$

This system can be written as Mx = b where M is a matrix and x and b are column vectors

```
>> M =[ 1 3 5; 2 1 1; 4 3 6]

M =
    1    3    5
    2    1    1
    4    3    6

>> b =[3; 2; 1 ]
b =
    3
    2
    1
>> x =M\b
x =
   -0.0909
    3.9091
   -1.7273
```

MATLAB can easily solve sets of linear equations if we place the coefficients of the terms appearing in those linear equations as matrices and vectors, as shown, and use the "backslash" operator \ to obtain the solution.

Determining the eigenvectors and eigenvalues of a matrix M
(For a real symmetric matrix the eigenvalues are real and the eigenvectors are real and orthogonal to each other)

[eigenvects, eigenvals] = eig(M)

```
>> [evects, evals] = eig(matrix)
evects =

  -0.8752    0.3507    0.3332
   0.4783    0.7300    0.4881
   0.0720   -0.5865    0.8067

evals =

   3.7531         0         0
        0   -1.0172         0
        0         0   11.2641
```

Eigenvectors (in columns)

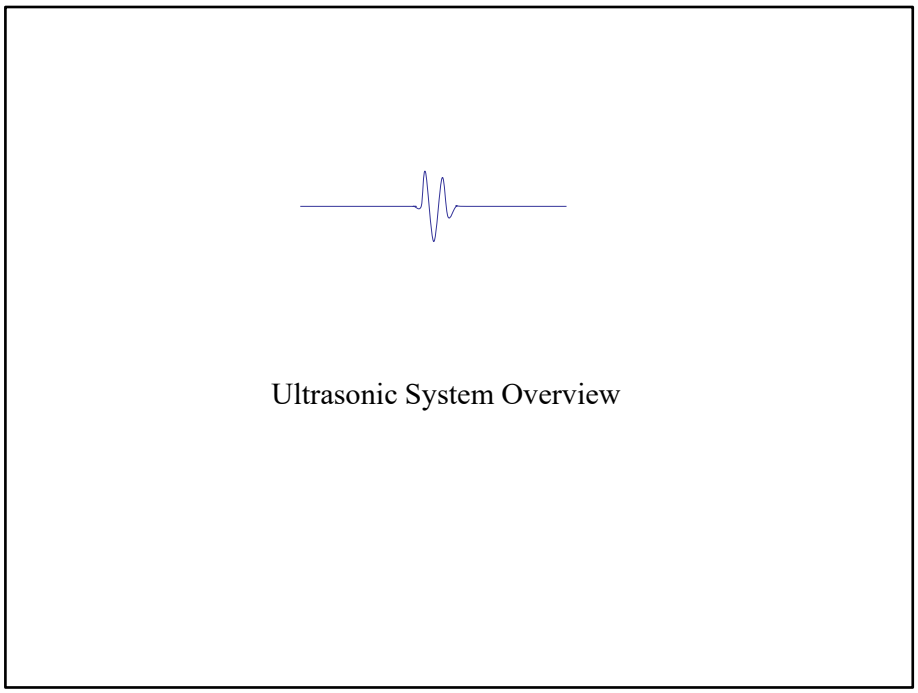
Corresponding eigenvalues

One important operation often performed on a matrix M is to determine its eigenvectors and eigenvalues. The MATLAB function eig performs the necessary calculations for us.

Homework Problems

M.1, M.2

Here are listed two homework questions that ask you to solve some problems that are associated with the waves involved in ultrasonic NDE inspections. See the home work manual for the actual questions and solutions.



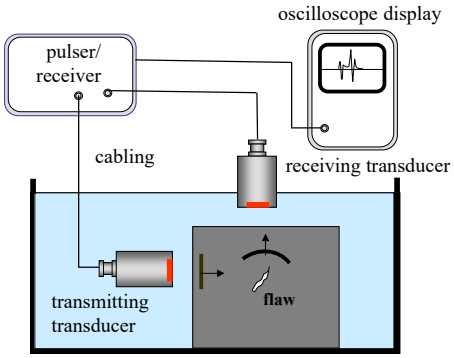
These slides are essentially Chapter 1 in the book: Schmerr, L.W. and S.J. Song, Fundamentals of Ultrasonic Evaluation Systems – Models and Measurements.

Learning Objectives

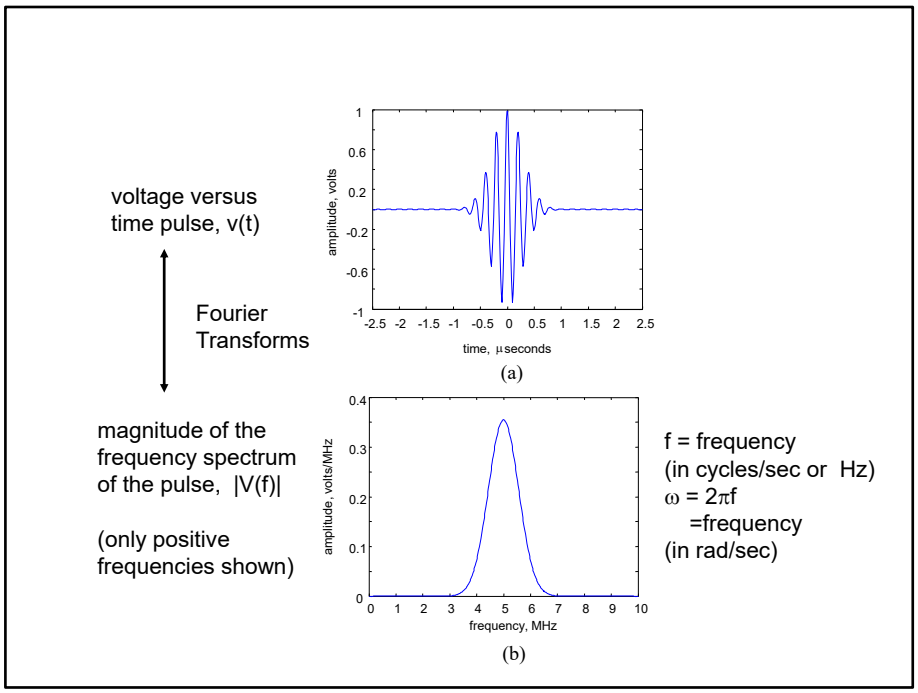
Overview of the UT system components that will be discussed

In subsequent presentations we are going to discuss all the elements that make up an ultrasonic NDE inspection system. This set of slides will walk through an entire system and give a preview of what is to come.

The ultrasonic system components

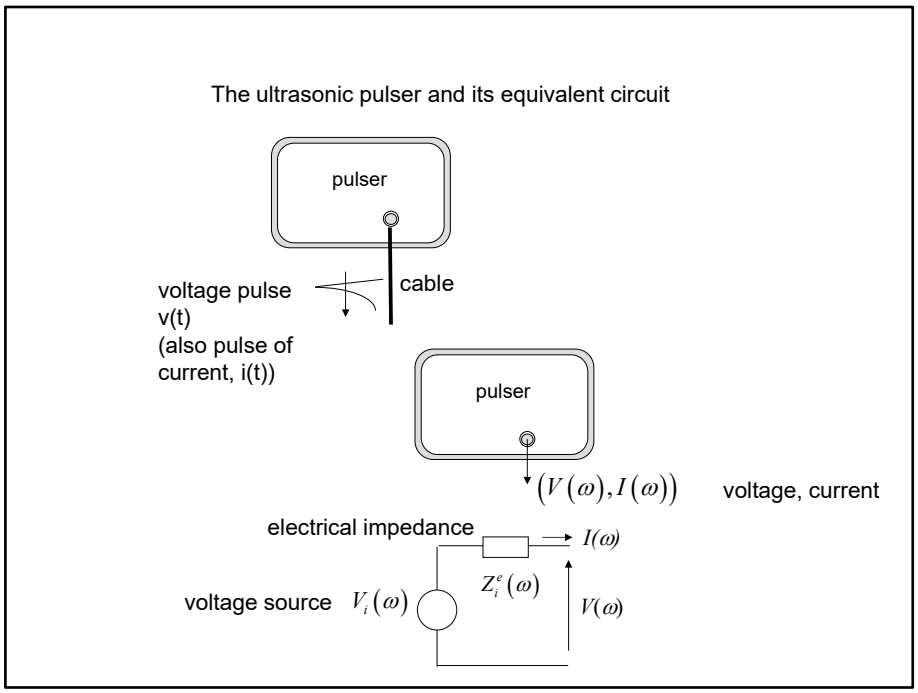


Here are all the components we will discuss in detail: the pulser/receiver, the cabling, the transducers, and the oscilloscope display. An immersion setup is shown only to illustrate these components.



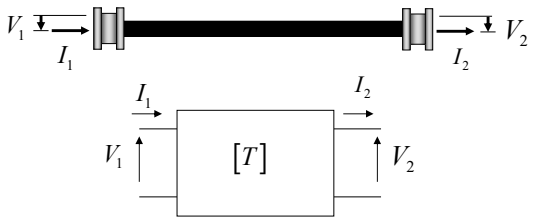
Although ultrasonic systems generate and receive pulses (short signals as a function of time) we can use a Fourier transform to generate the frequency components contained in the pulses and plot the so-called spectrum of the pulses as a function of frequency. Shown is an example of a short signal and its spectrum as might be seen in an ultrasonic NDE test. If we know the spectrum of the signal we can also use an inverse Fourier transform to recover the time signal.

Frequency is usually measured either in terms of Hertz (cycles/sec) or in terms of omega (radians/sec). In NDE tests, the very high frequencies involved are typically measured in millions of Hertz, or MHz, which stands for megaHertz



The pulser section of a pulser/receiver puts out repetitive short pulses of voltage and current as a function of time (only one is shown). If we compute the frequency components of these signals we can show that we can model how the pulser generates these outputs in the frequency domain by replacing the complex circuits contained in the pulser with a model that represents the pulser simply as a voltage source and an electrical impedance.

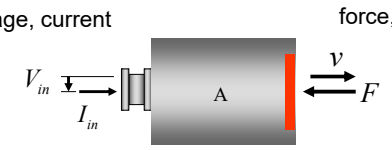
transmitting cabling and its equivalent transfer matrix



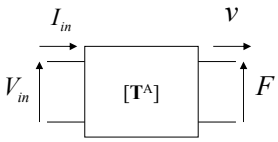
$$\begin{Bmatrix} V_1 \\ I_1 \end{Bmatrix} = \begin{bmatrix} T_{11}(\omega) & T_{12}(\omega) \\ T_{21}(\omega) & T_{22}(\omega) \end{bmatrix} \begin{Bmatrix} V_2 \\ I_2 \end{Bmatrix}$$

Normally we think of a cable as simply transferring signals from one location to another. However, at the MHz frequencies seen in ultrasonic NDE tests we will see this is not in general true unless the cable is short (roughly about a meter or less). In the frequency domain the cable can be modeled as a **two port system** where input voltage and current are related to the output voltage and current through a 2x2 **transfer matrix**. This transfer matrix can be determined though a series of electrical measurements.

a sending transducer and its equivalent transfer matrix



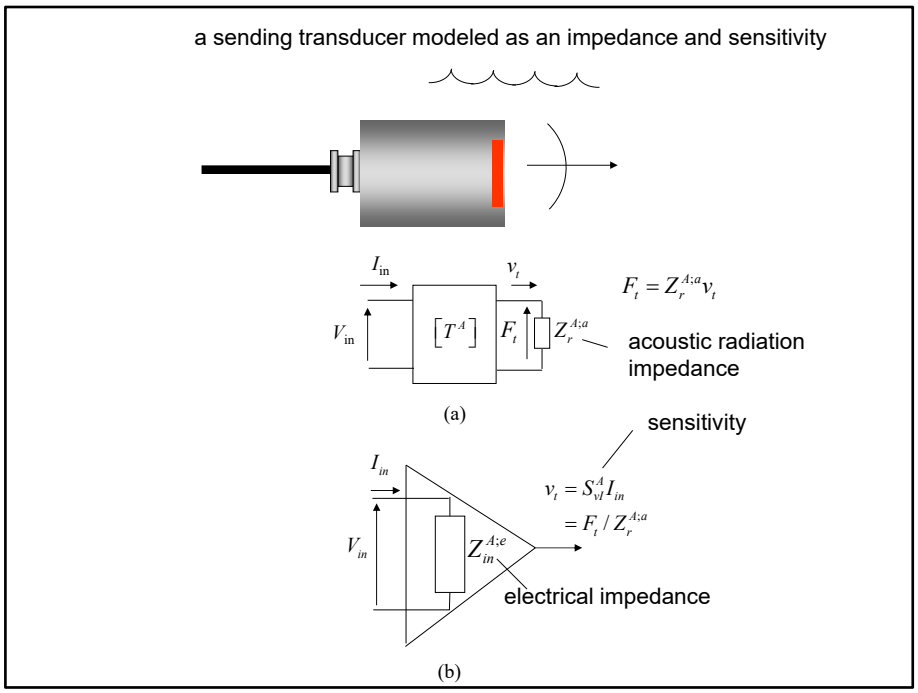
(a)



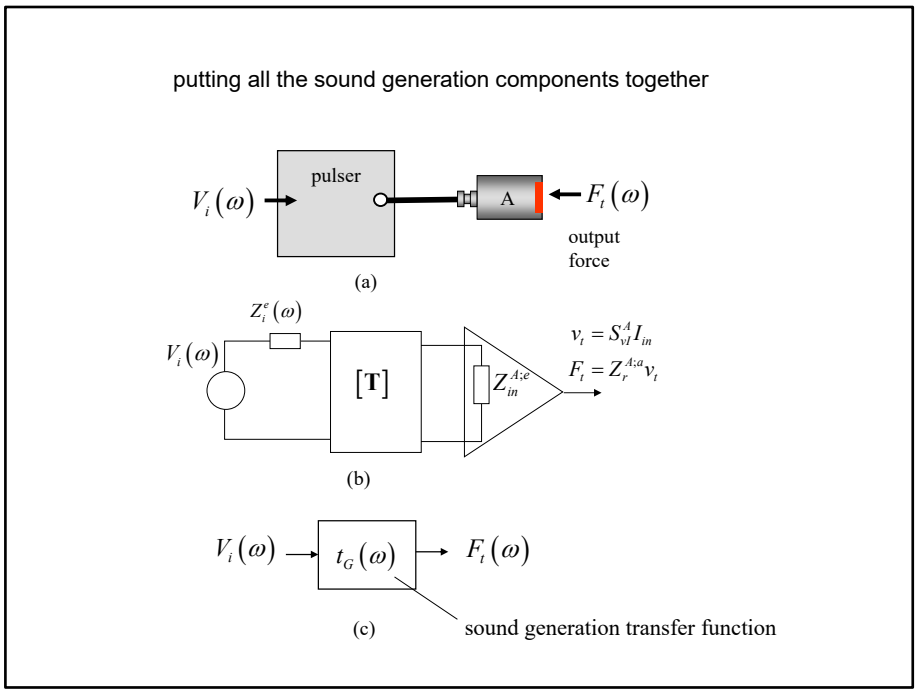
(b)

$$\begin{Bmatrix} V_{in} \\ I_{in} \end{Bmatrix} = \begin{bmatrix} T_{11}^A & T_{12}^A \\ T_{21}^A & T_{22}^A \end{bmatrix} \begin{Bmatrix} F \\ v \end{Bmatrix}$$

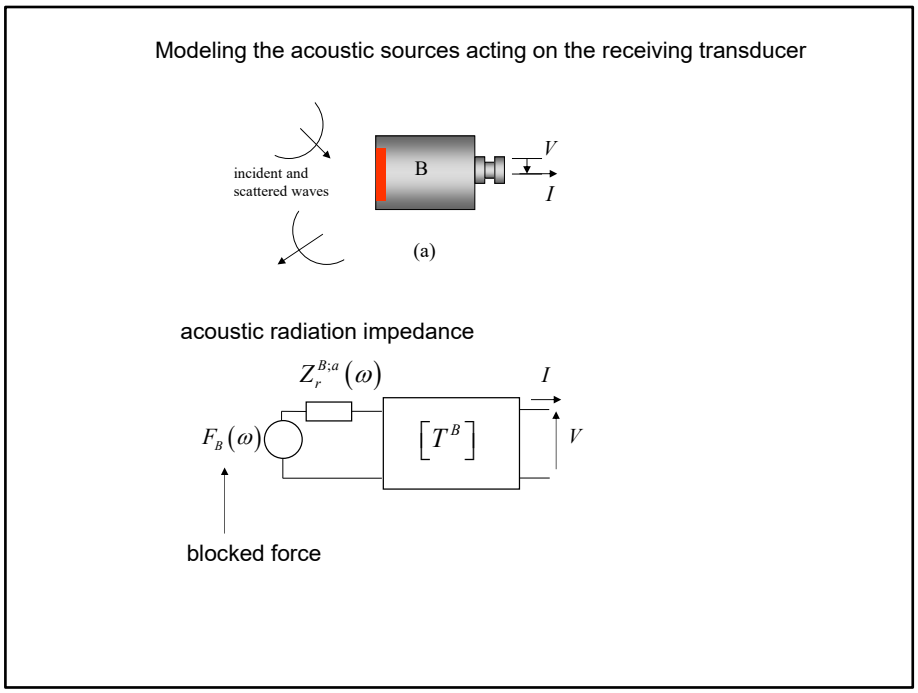
A transmitting transducer can also be modeled as a two port system where the voltage and current inputs are converted into mechanical outputs of force and velocity. Again, a 2x2 transfer matrix will characterize this two port system.



Unlike a cable, it is difficult to measure directly the transfer matrix components of a transducer with purely electrical measurements. However, because such a transducer is always used when the acoustic output side of the transducer is in contact with some medium (fluid or solid), the output force and velocity are always related to each other through an **acoustic radiation impedance**, i.e. the output of the transducer is mechanically terminated. Under these conditions we can replace the two port system with a transducer model consisting of an **electrical impedance** and a **sensitivity**, both of which can be determined with electrical measurements.

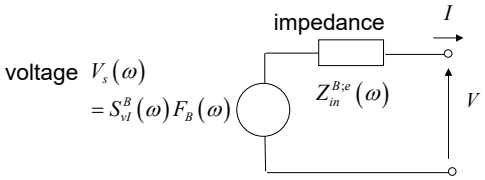


Models of all the elements contained in the sound generation process can be combined as shown. These elements can all be lumped into a single complex **sound generation transfer function** that relates the input voltage spectrum of the pulser to an acoustical output of the sending transducer such as the force. Thus, we can reduce the entire sound generation process into a model of single input-single output system characterized by the sound generation transfer function.



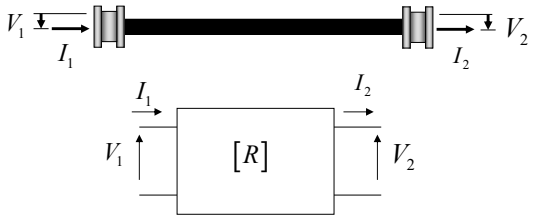
When a transducer is used a receiver, the waves incident on and scattered from the transducer act as acoustic driving terms as shown in (a). These waves can be modeled as a acoustic source, called the **blocked force** (which will be defined later) acting on the transducer face and the **acoustic radiation impedance** of the receiving transducer.

the receiving transducer also can be modeled by an impedance and sensitivity and these combined with the acoustic sources to yield a simple source and impedance model:



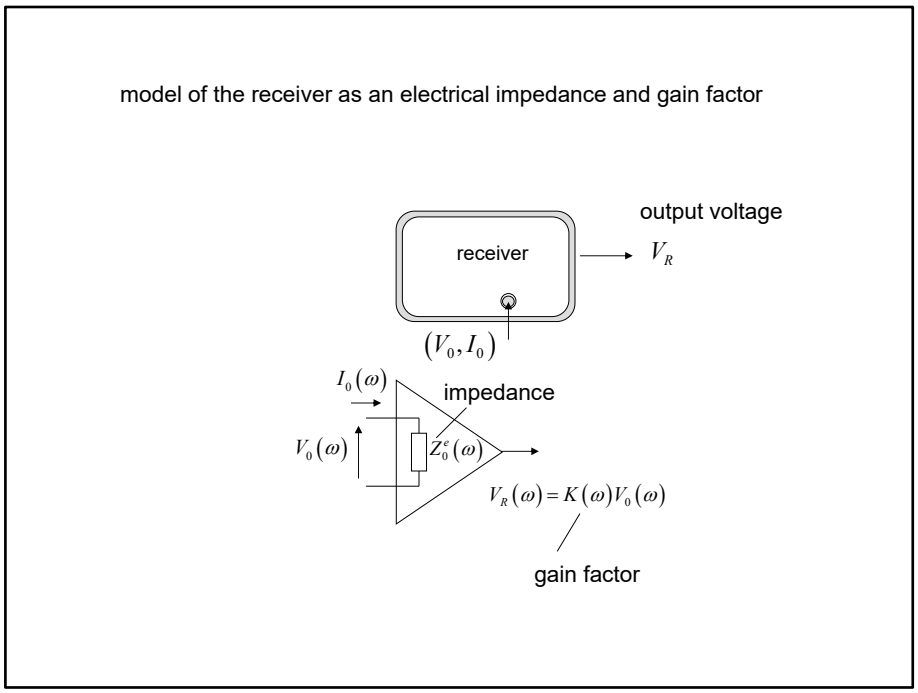
The 2x2 transfer matrix of the receiving transducer can also be replaced by an electrical impedance and sensitivity and these elements can be combined with the acoustic driving terms to model the receiving transducer and its acoustic sources by simply an equivalent voltage source and an electrical impedance.

receiving cabling and its equivalent transfer matrix

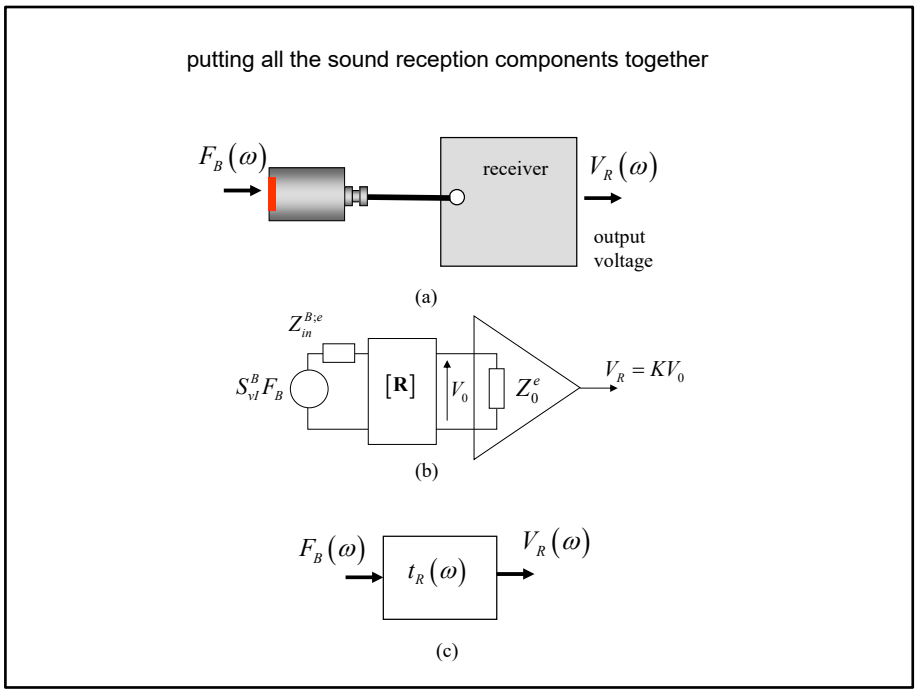


$$\begin{Bmatrix} V_1 \\ I_1 \end{Bmatrix} = \begin{bmatrix} R_{11}(\omega) & R_{12}(\omega) \\ R_{21}(\omega) & R_{22}(\omega) \end{bmatrix} \begin{Bmatrix} V_2 \\ I_2 \end{Bmatrix}$$

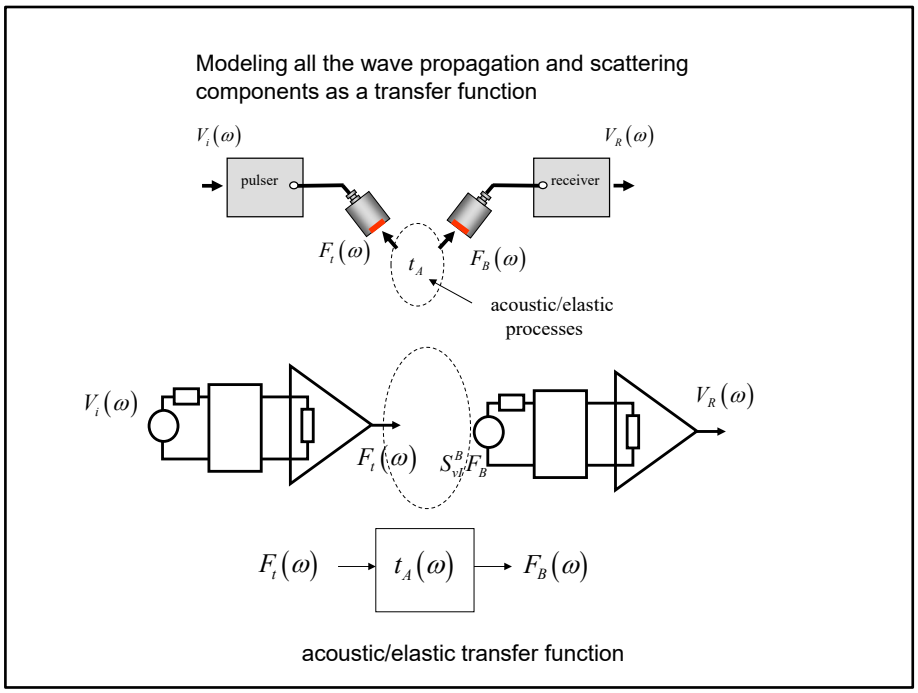
A cable between the receiving transducer and the receiver section of the pulser/receiver can again be modeled as a two port system characterized by a 2x2 transfer matrix.



The receiver section of a pulser/receiver acts to amplify the received signals and often provide low and high pass filters that modify the frequency content of the signals. In quantitative NDE studies we normally do not want to filter out any of the available information so that we will model the receiver as only an **electrical impedance** and a frequency dependent **gain factor**. If one wishes to add filters to this model, that is easily accomplished.

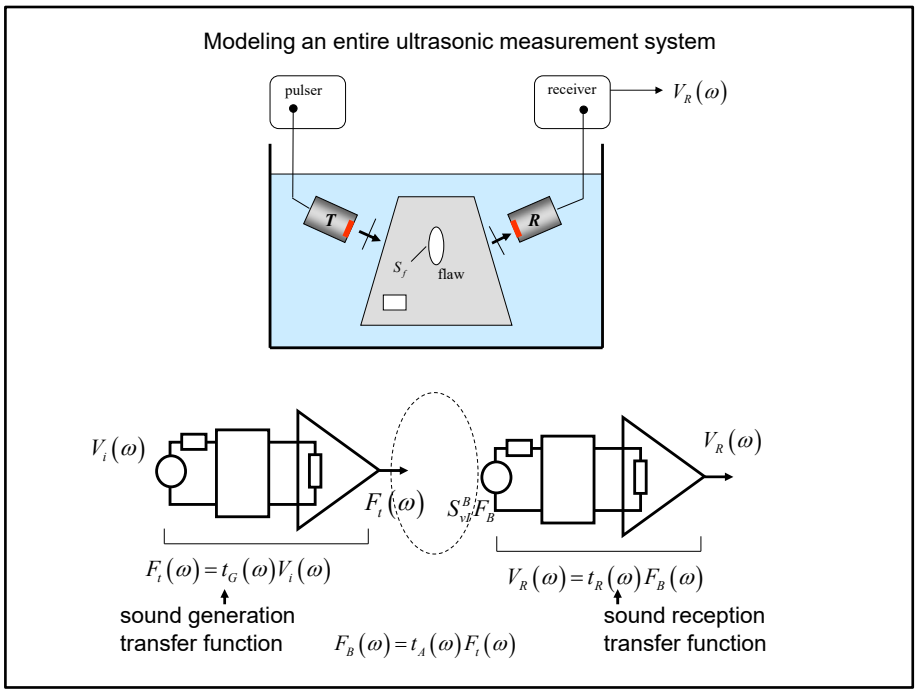


One can combine the receiving transducer, cable, and receiver into a composite model where all the elements are known. Again, these elements can be combined into a **receiving transfer function** that relates the blocked force to the received voltage in a single input, single output system that characterizes the receiving part of the entire system.



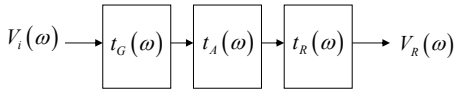
Since we have models of both the sound generation and sound reception parts of a measurement system, to complete the model for the entire system we need to characterize all the acoustic/elastic waves present between the transducers in terms of an **acoustic/elastic transfer function** that relates the force generated by the transmitting transducer to the blocked force acting on the receiving transducer.

Unlike the sending and receiving transfer functions, however, it is not possible to measure the acoustic/elastic transfer function since it due to complex 3-D wave interactions that are not directly accessible. Thus, one must model this transfer function with appropriate wave propagation and scattering models.

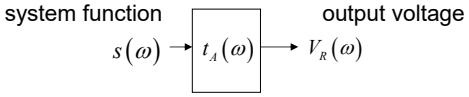


If we combine together all the elements we have described we now have a complete model of an ultrasonic NDE measurement system.

In terms of transfer functions the ultrasonic system looks like:



We can also combine the sound generation and reception transfer functions with the voltage source of the pulser tin a single function called the system function:



acoustic/elastic transfer function

$$V_R(\omega) = t_A(\omega)\, s(\omega)$$

A complete ultrasonic system model can thus be characterized in terms of the three transfer functions we have discussed. However, we can combine the sound generation and sound reception transfer functions together with the voltage source of the pulser into a single **system function** that we will show we can measure directly without knowing all the individual components that go into that system function. Thus, ultimately we can simply characterize a complete ultrasonic system as a single input, single output system where if we measure the system function and model the acoustic/elastic transfer function, we can predict the spectrum of the received voltage signals of the entire measurement system. An inverse Fourier transform of this spectrum then will produce the measure signals as a function of time.

The Fourier Transform

The following slides will give a brief introduction to the Fourier transform. It is this transform that allows us to characterize completely an entire ultrasonic system in terms of its components in a direct and simple manner.

Learning Objectives

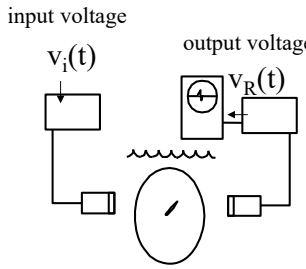
definition of Fourier Transform and its inverse
reasons for use of the Fourier Transform
brief introduction to Fourier transform properties
delta function and its spectrum
definition of dB scale
definition of -6dB bandwidth
bandwidth - time domain connection

We will define the Fourier transform and its inverse and give the reasons why this transform is so useful. We will also describe some of the characteristics of Fourier transforms and give an example of a very important special case – the Fourier transform of a delta function.

We will define the decibel scale often used for ultrasonic signals and the characterization of the frequency domain spectrum of a signal in terms of its -6 dB bandwidth.

Finally, we will demonstrate the connection between the characteristics of a signal in the time domain and in the frequency domain.

Fourier Transforms for UT Systems



Time domain ↔ Frequency domain

Fourier transform

$$V(\omega) = \int_{-\infty}^{+\infty} v(t)\exp(i\omega t)\,dt$$

$\omega \ldots$ rad/sec

inverse Fourier transform

$$v(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} V(\omega)\exp(-i\omega t)\,d\omega$$

Even though an ultrasonic measurement system inherently involves time-dependent signals such as the input voltage of the pulser or the measured output voltage at the receiver, as shown, it is easier, as stated earlier, to characterize all the elements of the measurement system in the frequency domain. Fortunately, the **Fourier transform** allows us to generate the frequency domain signals from those in the time domain. Similarly, the **inverse Fourier transform** allows us to reverse this process and generate the time domain signals from those in the frequency domain.

Equivalent forms

$$V(\omega) = \int_{-\infty}^{+\infty} v(t)\exp(i\omega t)dt$$

$$v(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} V(\omega)\exp(-i\omega t)d\omega$$

or $\omega = \left(2\pi \frac{rad}{cycle}\right)\left(f \frac{cycle}{\text{sec}}\right)$

$$V(f) = \int_{-\infty}^{+\infty} v(t)\exp(2\pi i f t)dt$$

$$v(t) = \int_{-\infty}^{+\infty} V(f)\exp(-2\pi i f t)df$$

$V(\omega)$ *or* $V(f)$ … dimensions are volts-sec or volts-μsec

for NDE problems t is usually in μsec, f in MHz

Shown are the definitions we will use of the Fourier transform and its inverse. Those transforms can be written in terms of the frequency in rad/sec or in terms of cycles/sec (Hz). In NDE tests the time signals are often very short so they are typically measured in microseconds (μsec) and the frequencies are very high so that the corresponding frequency domain signals are typically measured in terms of millions of cycles/sec (MHz).

Note that if v(t) is a time domain voltage signal with t measured in seconds, its frequency spectrum V(f) has dimensions of volts-sec, but because we will often work entirely in the frequency domain, we will ignore this difference and refer to the frequency spectrum signals as simply volts. Other quantities such as pressure, velocity, current, etc. will follow this naming convention.

Fourier Transforms

$$V(f) = \int_{-\infty}^{+\infty} v(t) \exp(2\pi i f t)\, dt$$

$$v(t) = \int_{-\infty}^{+\infty} V(f) \exp(-2\pi i f t)\, df$$

A few properties of Fourier Transforms

If $v(t) \leftrightarrow V(f)$

then $v(t - t_0) \leftrightarrow \exp(2\pi i f t_0) V(f)$

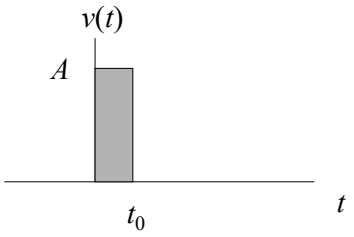
$$\frac{dv}{dt} \leftrightarrow -2\pi i f\, V(f)$$

Because of the complex exponential term in the Fourier transform, the frequency domain signal V(f) will usually be complex even though the time domain signal v(t) is real.

Using this definition it is easy to derive a number of relationships between time domain signals and frequency domain signals. Two important cases are shown here where we see that a shifting (delay) of a time signal causes the frequency spectrum to be multiplied by a complex exponential term. Similarly, differentiating a time domain signal is equivalent to multiplication by an imaginary frequency dependent term in the frequency domain.

Example Fourier transform

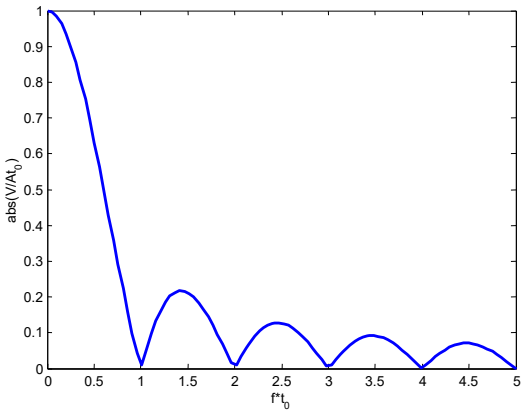
$$V(f) = \int_{-\infty}^{+\infty} v(t) \exp(2\pi i f t)\, dt$$



$$\begin{aligned} V(f) &= \int_0^{t_0} A \exp(2\pi i f t)\, dt \\ &= \frac{A \exp(2\pi i f t)}{2\pi i f}\Bigg|_0^{t_o} \\ &= \frac{A\left[\exp(2\pi i f t_0) - 1\right]}{2\pi i f} \\ &= \frac{A t_0 \exp(i\pi f t_o)\sin(\pi f t_o)}{\pi f t_0} \end{aligned}$$

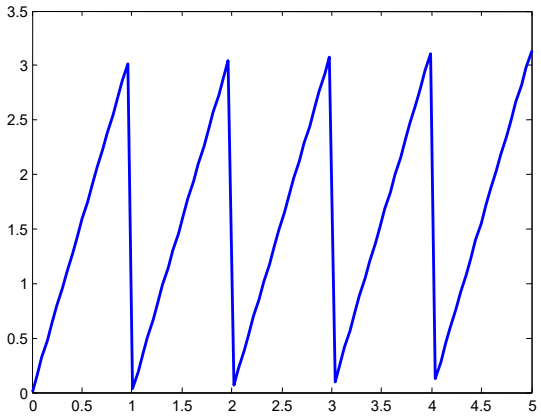
As a simple example of a Fourier transform, consider the box function shown in the time domain. In this case the Fourier transform is easy to calculate and we see that indeed this the transform is a complex function in the frequency domain.

```
>> u=linspace(0, 5, 100);
>> u = u + eps*(u = = 0);
>> yf = exp(i*pi*u).*sin(pi*u)./(pi*u);
>> plot(u, abs(yf))
>> xlabel('f*t_0')
>> ylabel('abs(V/At_0)')
```

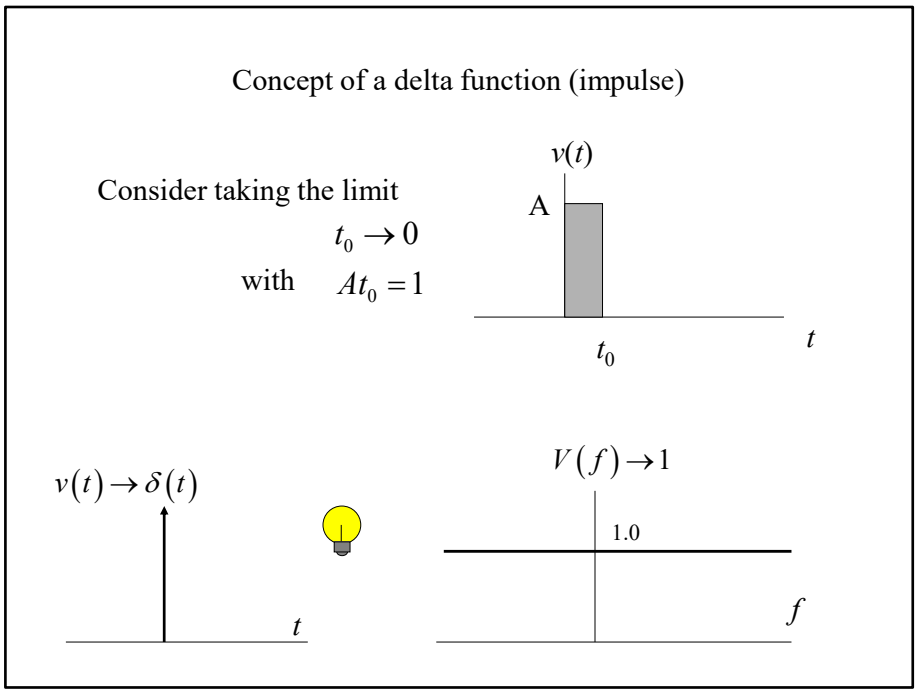


If we evaluate the Fourier transform of the box function and plot its normalized magnitude versus non-dimensional frequency for positive frequencies we see most of the frequency content is in a low frequency “lobe” with decreasing content in side lobes as the frequency increases.

```
>> plot(u, angle(yf))
```



If we plot the phase of the function we see a general linear increasing behavior with jumps of pi radians. The linear behavior comes from the complex exponential term in the function and the jumps of pi radians occur because of the periodic changes in sign of the sine function (recall exp(i π) = -1).

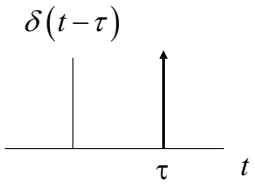


If we imagine taking the limit of the box function as its time duration goes to zero but where we keep the product of its amplitude and duration at unity, then conceptually we obtain an infinite spike of zero duration at time t = 0, which we call a **delta function**. In the frequency domain the spectrum simply goes to a constant value of one at all frequencies. Thus, a delta function excites all frequencies equally which makes it an ideal input function to characterize a system's frequency domain response.

We will indicate that the delta function is a **key concept** by placing a **light bulb** next to it. In later slides we will use the light bulb symbol to highlight other key concepts.

Some properties of delta functions

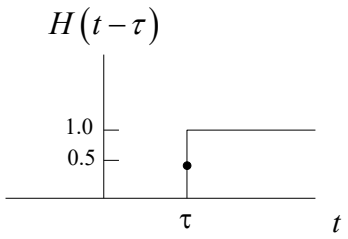
$$\int_a^b g(t)\,\delta(t-\tau)\,dt = \begin{cases} 0 & \tau < a \text{ or } \tau > b \\ g(\tau) & a < \tau < b \\ g(\tau)/2 & \tau = a \text{ or } \tau = b \end{cases}$$



$$\delta(t-\tau) = 0 \quad t \neq \tau$$

$$\int_{u=-\infty}^{u=t} \delta(u-\tau)\,du = \begin{cases} 0 & t < \tau \\ 1/2 & t = \tau \\ 1 & t > \tau \end{cases}$$

$$= H(t-\tau) \quad \text{unit step function}$$

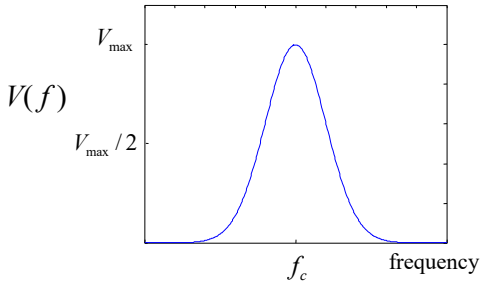


Here we list several important properties of delta functions. We see that a delta function multiplied by an "ordinary" function in an integral has the property of sampling that ordinary function at the location of the delta function. Similarly, we see that an integral of a delta function generates a **unit step function** at the location of the delta function.

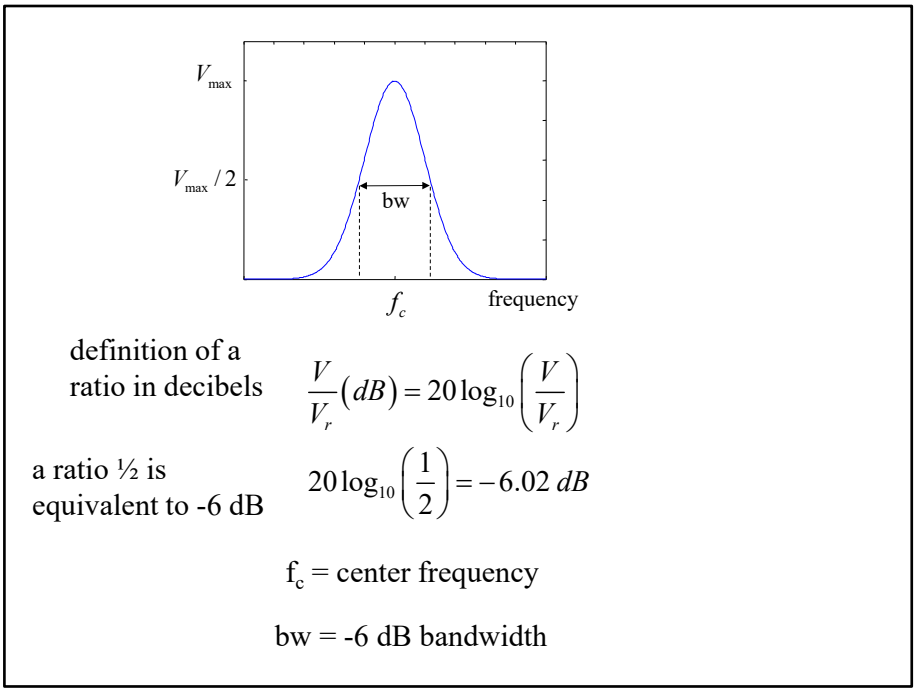
$$v(t) = \cos(2\pi f_c t)\exp\left[-\pi t^2 / 4A^2\right]$$

produces a Gaussian spectrum with a center frequency f_c and bandwidth bw where $A = \dfrac{\sqrt{\ln 2}}{\pi\, bw}$

$$V(f) = \sqrt{\pi} A\left\{\exp\left[-4\pi^2 A^2 (f - f_c)^2\right] + \exp\left[-4\pi^2 A^2 (f + f_c)^2\right]\right\}$$



This slide shows an example time domain signal that generates a spectrum consisting of two real Gaussians, one centered at a positive frequency and one centered at a negative frequency. Only the Gaussian centered at the positive frequency is shown in the plot.



definition of a ratio in decibels

$$\frac{V}{V_r}(dB) = 20\log_{10}\left(\frac{V}{V_r}\right)$$

a ratio ½ is equivalent to -6 dB

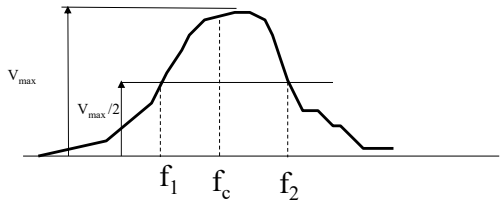
$$20\log_{10}\left(\frac{1}{2}\right) = -6.02\ dB$$

f_c = center frequency

bw = -6 dB bandwidth

Gaussians are often used to discuss signal characteristics. For example, we can define the bandwidth of the Gaussian as the width of the function, bw, where the function drops to one half of its maximum value. The value of ½, as measured in decibels (dB) is -6 dB so this is also called the **-6 dB bandwidth**. Similarly, we can call the frequency at the location of the maximum of the Gaussian as the **center frequency** of the signal.

In real ultrasonic systems the spectrum is not symmetric, so how do we define a bandwidth and center frequency?



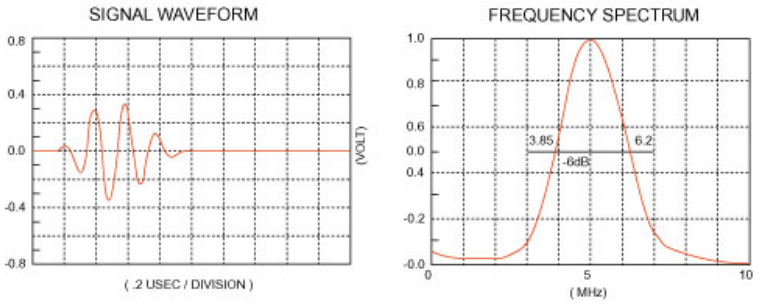
center frequency: $f_c = \frac{f_1 + f_2}{2}$

-6 dB bandwidth (in % of center frequency):

$$bw = \frac{f_2 - f_1}{f_c} \times 100$$

In a real ultrasonic signal (whose spectrum is not symmetric like the Gaussian) we can determine the maximum magnitude of the spectrum and then find the frequencies where the signal is one half of that maximum value. Those two frequencies then can be used to determine both a center frequency and -6dB bandwidth for the signal as shown.

Transducer specifications (from the manufacturer)



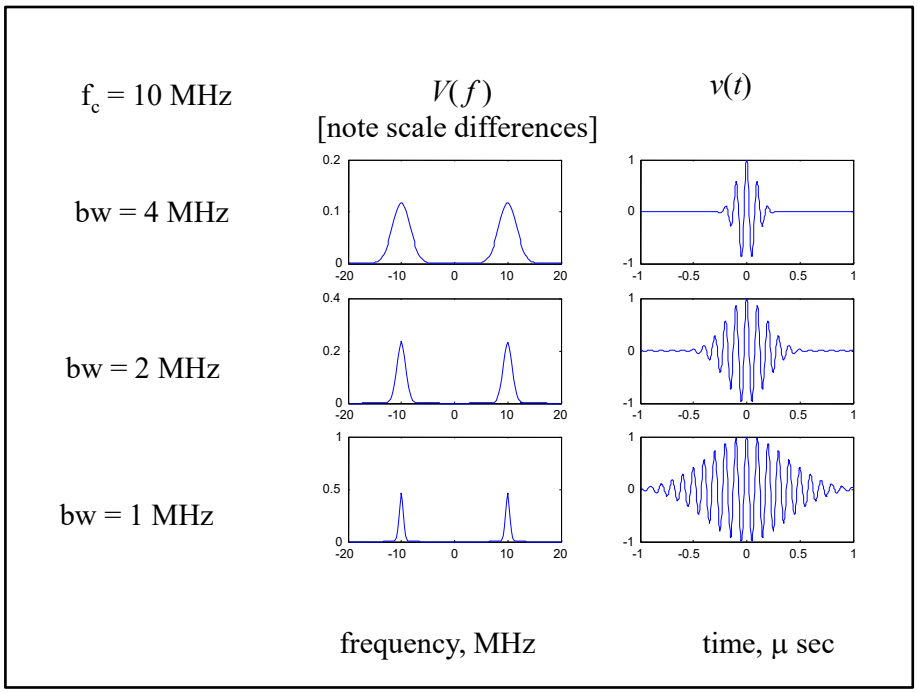
MEASUREMENTS PER ASTM E1065

WAVEFORM DURATION:
−14DB LEVEL −− .546US
−20DB LEVEL −− .650US
−40DB LEVEL −− 1.24US

SPECTRUM MEASURANDS:
CENTER FREQ. −−−− 5.03MHz
PEAK FREQUENCY −− 4.95MHz
−6DB BANDWIDTH −− 46.77 %

- 6 dB bandwidth in % of center frequency

Here is an example of a specification sheet obtained from a transducer manufacturer that shows plots of the time domain and frequency domain outputs of the transducer and where we see they give both the center frequency and -6 dB bandwidth.



We can use the time domain signal discussed previously that generates Gaussians in the frequency domain to examine how changes in the time domain and frequency domain are related. Here, we keep the center frequency fixed at 10 MHz and change the bandwidth. We plot both positive and negative frequencies to show that there are actually two Gaussians present. At the largest bandwidth we see that the time domain signal is the shortest, becoming wider in the time domain as the bandwidth becomes smaller. Note that the amplitude of the frequency domain signals also gets larger as the bandwidth decreases even though the amplitude of the time domain signal is unchanged. This behavior may not be as directly evident since we see the scale of the response in the frequency domain is changed for the various cases to make the plots more readable.

```
% Gaussian_script
f= linspace(-20, 20, 200);
t =linspace(-1,1, 500);
subplot(3,2,1)
bw =4;
fc = 10;
[y,z]= Gauss_funcs(f,t,fc, bw);
plot(f, y)
subplot(3, 2, 2)
plot(t,z)
subplot(3, 2, 3)
bw =2;
fc = 10;
[y,z]= Gauss_funcs(f,t,fc, bw);
plot(f,y)
subplot(3,2,4)
plot(t,z)
subplot(3, 2, 5)
bw =1;
fc = 10;
[y,z]= Gauss_funcs(f,t,fc, bw);
plot(f,y)
subplot(3,2,6)
plot(t,z)
```

```
function [y, z] =Gauss_funcs(f,t,fc,bw)
a = sqrt(log(2))/(pi*bw);
y = sqrt(pi)*a*(exp(-(2*a*pi*(f - fc)).^2) + exp(-(2*a*pi*(f + fc)).^2));
z = cos(2*pi*fc*t).*exp(-(1/(4*a^2))*t.^2);
```

Here is the MATLAB code that generates the plots seen on the previous slide. The Matlab m-files for all the functions and scripts discussed in this course can be downloaded from the course website.

Note: there are a number of forms used for the Fourier Transforms. Some of these are different from the one we will use here exclusively:

$$V(\omega) = \int_{-\infty}^{+\infty} v(t) \exp(i\omega t) dt$$

$$v(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} V(\omega) \exp(-i\omega t) d\omega$$

Some other examples:

$$V(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} v(t) \exp(i\omega t) dt$$

$$v(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} V(\omega) \exp(-i\omega t) d\omega$$

(often seen in the math literature)

Our definition of the Fourier transform and its inverse are not the only ones seen in the literature. Here, and on the next slide are some examples of different definitions.

$$V(\omega) = \int_{-\infty}^{+\infty} v(t)\exp(-j\omega t)\,dt \qquad j = \sqrt{-1}$$

(often seen in the EE literature)

$$v(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} V(\omega)\exp(+j\omega t)\,d\omega$$

All of these forms are acceptable. In fact we could write Fourier transform pairs in general as:

$$V(\omega) = N_1 \int_{-\infty}^{+\infty} v(t)\exp(\pm i\omega t)\,dt$$

$$v(t) = N_2 \int_{-\infty}^{+\infty} V(\omega)\exp(\mp i\omega t)\,d\omega$$

as long as $N_1 N_2 = \frac{1}{2\pi}$

All of these definitions are acceptable but we will only use our original choice in all subsequent work to be consistent. Choosing these other definitions will lead to differences in amplitude and/or phase from our definition so you must be careful when comparing our results with others that may appear in the literature.

References

Sneddon, I.N., **Fourier Transforms**, McGraw-Hill, New York, 1951.

Bracewell, R.N., The **Fourier Transform and its Applications**, **3rd Ed**. McGraw-Hill, New York, 2000.

Here are several basic references on Fourier transforms. There are many more.

The Fast Fourier Transform (FFT)
and MATLAB Examples

In practice, Fourier transforms are calculated with a specific numerical procedure called a **fast Fourier transform** (FFT) so these slides will describe in detail how those calculations are performed, using MATLAB to illustrate specific examples.

Learning Objectives

Discrete Fourier transforms (DFTs) and their relationship to the Fourier transforms

Implementation issues with the DFT via the FFT

- sampling issues (Nyquist criterion)
- resolution in the frequency domain (zero padding)
- neglect of negative frequency components

We will examine discrete Fourier transforms and their connections to the Fourier transform and show how discrete Fourier transforms are calculated with the fast Fourier transform algorithm. Various FFT implementation issues will be discussed.

$$V(f) = \int_{-\infty}^{+\infty} v(t) \exp(2\pi i f t) dt$$

$$v(t) = \int_{-\infty}^{+\infty} V(f) \exp(-2\pi i f t) df$$

These Fourier integral pairs normally are performed numerically by sampling the time and frequency domain functions at discrete values of time and frequency and then using the discrete Fourier transforms to relate the sampled values

If ultrasonic signals are sampled, then instead of performing these Fourier integrals we need to perform sums of sampled values. These sums define the discrete Fourier transforms.

Fast Fourier Transform

Discrete Fourier Transform

$$V_p(f_n) = \frac{T}{N} \sum_{j=0}^{N-1} v_p(t_j) \exp(2\pi i jn / N)$$

$$v_p(t_k) = \frac{1}{T} \sum_{n=0}^{N-1} V_p(f_n) \exp(-2\pi i kn / N)$$

$T = N\Delta t$ … time window sampled with N points

Δt … sampling time interval

$t_j = j\Delta t \qquad f_n = n\Delta f = n / N\Delta t$

These are the discrete Fourier transform and its inverse which arise from sampling the Fourier integrals we defined earlier.

Fast Fourier Transform

As with the Fourier transforms, there are different choices made for the Discrete Fourier transform pairs. In general, we could write:

$$V_p\left(f_n\right) = n_1 \sum_{j=0}^{N-1} v_p\left(t_j\right) \exp\left(\pm 2\pi i\, jn / N\right)$$

$$v_p\left(t_k\right) = n_2 \sum_{n=0}^{N-1} V_p\left(f_n\right) \exp\left(\mp 2\pi i\, kn / N\right)$$

as long as $n_1 n_2 = \frac{1}{N}$

The indexing could also go from 1 to N instead of 0 to N-1

The exact form of the discrete Fourier transform pairs, like the Fourier transform integrals, depend on the specific form we choose for those transforms. Note that these represent periodic functions while our ultrasonic signals are typically aperiodic. We have placed a p subscript on these quantities to indicate this periodicity. We will see that we must satisfy certain conditions to ensure this periodicity does not affect the calculations.

Fast Fourier Transform

$$V_p(f_n) = \frac{T}{N}\sum_{j=0}^{N-1} v_p(t_j)\exp(2\pi i jn / N)$$

$$v_p(t_k) = \frac{1}{T}\sum_{n=0}^{N-1} V_p(f_n)\exp(-2\pi ikn / N)$$

These discrete Fourier Transforms can be implemented rapidly with the Fast Fourier Transform (FFT) algorithm

	number of multiplications: direct calculation	number of multiplications: with the FFT
N	$(N-1)^2$	$(N/2)\log_2 N$
256	65,025	1,024
1,024	1,046,529	5,120
4,096	16,769,025	24,576

FFTs are most efficient if the number of samples, N, is a power of 2. Some FFT software implementations require this.

Performing these discrete Fourier transforms directly is computationally intensive but if we use a particular algorithm, called the fast Fourier transform (FFT) the number of multiplications can be drastically reduced as shown in the table. We will not describe the details of an FFT algorithm but there are many references available that do give that information. Note that some FFT algorithms require the number of sampled values be a power of 2 since that choice typically makes the algorithm more efficient.

Fast Fourier Transform

Mathematica

Fourier[{a_1,a_2,…,a_N}] ↕ lists InverseFourier[{b_1,b_2,…,b_N}]

$$\frac{1}{\sqrt{N}}\sum_{r=1}^{N} a_r \exp\left[2\pi i(r-1)(s-1)/N\right]$$

$$\frac{1}{\sqrt{N}}\sum_{s=1}^{N} b_s \exp\left[-2\pi i(r-1)(s-1)/N\right]$$

Maple

FFT(N, x_{re}, x_{im}) ↕ arrays iFFT(N,X_{re},X_{im})

$$\sum_{j=0}^{N-1} x(j)\exp\left[-2\pi i jk/N\right]$$

$$\frac{1}{N}\sum_{k=0}^{N-1} X(k)\exp\left[2\pi i jk/N\right]$$

$N = 2^n$

MATLAB

fft(x) ↕ arrays ifft(X)

$$\sum_{j=1}^{N} x(j)\exp\left[-2\pi i(j-1)(k-1)/N\right]$$

$$\frac{1}{N}\sum_{j=1}^{N} X(j)\exp\left[2\pi i(j-1)(k-1)/N\right]$$

Different software packages implement the FFT in different forms. Here are some of the common cases.

Fast Fourier Transform

FFT function

```
function y = FourierT(x, dt)
% FourierT(x,dt) computes forward FFT of x with sampling time interval dt
% FourierT approximates the Fourier transform  where the integrand of the
% transform is x*exp(2*pi*i*f*t)
% For NDE applications the frequency components are normally in MHz,
% dt in microseconds
[nr, nc] = size(x);
if nr == 1
   N = nc;
else
   N = nr;
end
y = N*dt*ifft(x);
```

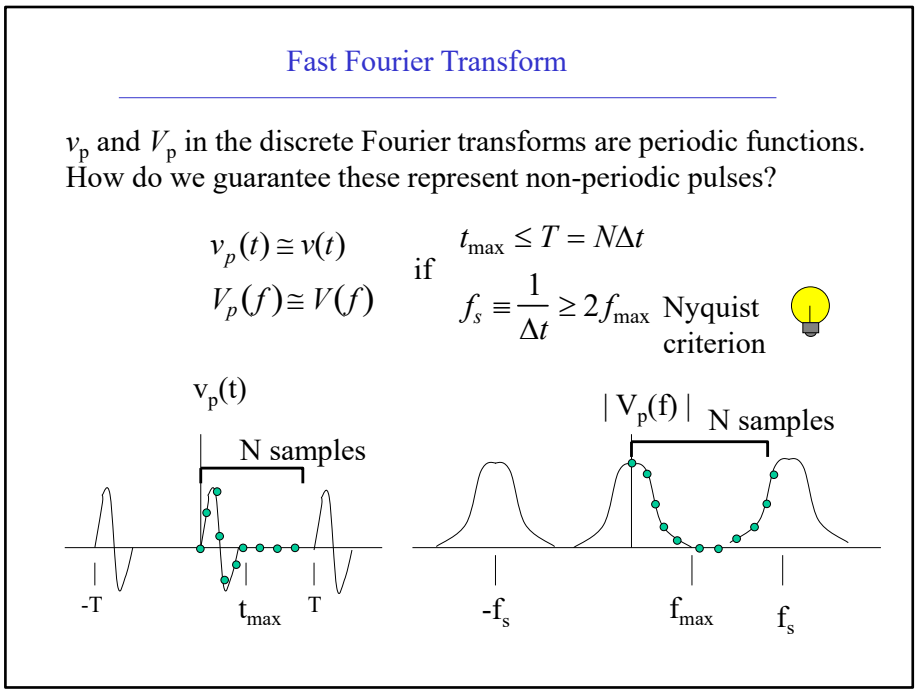
We will use the MATLAB functions (fft and ifft) to define new functions (FourierT and IFourierT) that are compatible with our definitions of the Fourier and discrete Fourier transforms and there inverses. Here is our FFT function FourierT.

Fast Fourier Transform

inverse FFT function

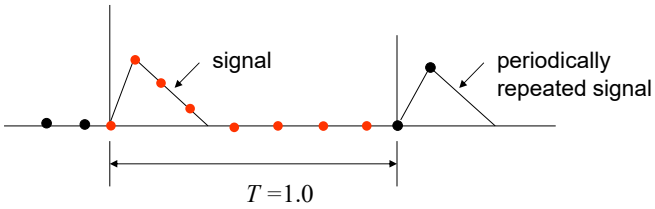
```
function y = IFourierT(x, dt)
% IFourierT(x,dt) computes the inverse FFT of x, for a sampling time interval dt
% IFourierT assumes the integrand of the inverse transform is given by
% x*exp(-2*pi*i*f*t)
% The first half of the sampled values of x are the spectral components for
% positive frequencies ranging from 0 to the Nyquist frequency 1/(2*dt)
% The second half of the sampled values are the spectral components for
% the corresponding negative frequencies. If these negative frequency
% values are set equal to zero then to recover the inverse FFT of x we must
% replace x(1) by x(1)/2 and then compute 2*real(IFourierT(x,dt))

[nr,nc] = size(x);
if nr == 1
        N = nc;
else
        N = nr;
end
y =(1/(N*dt))*fft(x);
```

Here is our inverse FFT function IFourierT



Shown here is a typical transient (aperiodic) ultrasonic signal and its spectrum, both repeated periodically. As long as the total time, T, of the sample window exceeds the time duration of the signal, and the sampling frequency is twice the maximum frequency present in the spectrum of the signal, then there is no overlap and we will obtain faithful representation of our signals in both the time and frequency domain, as shown. The condition that must be met on the sampling frequency is called the **Nyquist criterion**. In ultrasonic tests the frequencies involved typically are less than 20 MHz because of the transducers used and that fact that material attenuation also removes higher frequencies so one often samples ultrasonic NDE signals at a 100MHz sampling frequency, which is a very conservative choice satisfying the Nyquist criterion. Note that in the frequency domain the sampled spectrum values at negative frequencies are contained in the upper half of the sampled frequency domain values.

The use of linspace and s_space functions for FFTs and inverse FFTs



$$\Delta t = T / N$$

In the example above $N = 8$, $T = 1.0$ so $\Delta t = 1/8 = 0.125$

```
>> t = linspace(0,1, 8);
>> dt = t(2) -t(1)
dt =    0.1429
```

In modeling our sampled ultrasonic signals we cannot use the built-in MATLAB function linspace as it does not generate the proper samples. To illustrate, consider this simple time domain function which is sampled over a time window T=1 with N = 8 samples so we want the sampling interval to be 1/8 and the sampled values to be the ones shown as red dots. If we use linspace(0, 1, 8) we see an incorrect sampling interval (and the actual time values (not shown) will be incorrect)

To get the proper sampled values and the correct Δt we must use the alternate function s_space (defined later)

```
>> t = s_space(0,1, 8);
>> dt = t(2)-t(1)
dt = 0.1250
```

```
>> t = linspace(0,1, 512);
>> dt = t(2) - t(1)
dt = 0.0020
```

```
>> t = s_space(0,1, 512);
>> dt = t(2) - t(1)
dt = 0.0020
```

We have defined a new MATLAB function, **s_space**, that generates the proper values, as shown above for our previous example. Note that if the number of sample points is large, then these differences between linspace and s_space are not as apparent, as we can see if we make N = 512. However, those differences are still present (see the next slide) so in generating sampled time and frequency domain functions we will always use s_space rather than linspace.

When you are using many points the difference is not large if we use either linspace or s_space but the difference is still there

```
>> format    long
>> t = linspace(0, 1, 500);
>> dt = t(2) -t(1)

dt = 0.00200400801603

>> t = s_space(0, 1, 500);
>> dt = t(2) -t(1)

dt = 0.00200000000000
```

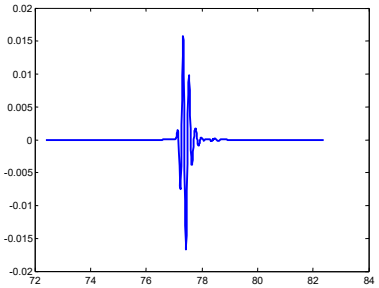
If we use the Matlab command format long we can see the small differences between the use of s_space and linspace

```
function  y = s_space(xstart, xend, num)
% S_SPACE(XSTART,XEND, NUM) generates num evenly spaced sampled
% values from xstart to (xend - dx), where dx is the sample
% spacing. This is useful in FFT analysis where we generate
% sampled periodic functions. Example: generate 1000
% sampled frequencies from 0 to 100MHz via f =s_space(0,100,1000);
% In this case the last value of f will be 99.9 MHz and the
% sampling interval will be 100/1000 =0.1 MHz.


ye =linspace(xstart, xend, num+1);

y=ye(1:num);
```

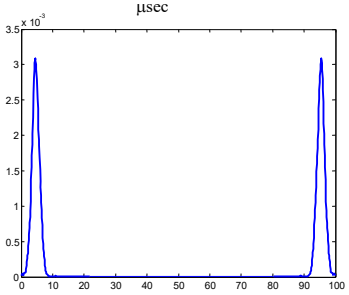
Here is the **s-space function** which we see is a simple modification of the use of linspace to get the proper sampled values and sampling interval.

Example of use of FourierT on a real ultrasonic signal

```
>> load FBH_flaw_n8      flat-bottom hole
>> size(texp)            signal

ans =
      1      1000        time values

>> size(vexp)

ans =
      1      1000        voltage values
>> plot(texp, vexp)
>> dt = texp(2) - texp(1)

dt =

   0.0100                sampling interval

>> fs = 1/dt

fs =

  100.0000               sampling frequency

>> vf = FourierT(vexp, dt);
>> f = s_space(0, fs, 1000);
>> plot(f, abs(vf))
```



µsec



MHz

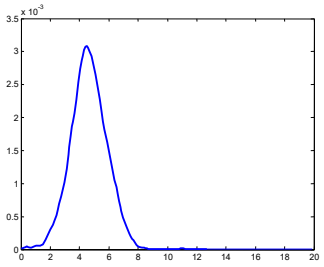
Here is an example of Fourier analysis of an actual ultrasonic signal, which in this case is the signal reflected from a flat-bottom hole in a metal sample(a common reference reflector used in NDE).

The sampled voltage versus time signal is stored in a vector, vexp, and the sampled times are in the vector texp, both of which are loaded from a saved file into MATLAB. If we examine those vectors we see there are 1000 sampled time and voltage values and the sampling interval is 0.01 (microseconds) and the sampling frequency is 100 (MHz). The received time domain waveform is plotted.

If we do the FFT with our MATLAB function FourierT, we must generate a proper set of 1000 frequencies with s-space, after which we can plot the magnitude of the spectrum. We see most of the spectrum is between zero and 10 MHz and we see the corresponding negative frequency components in the upper half of the spectrum. Clearly our sampling frequency was more than sufficient in this case to satisfy the Nyquist criterion.

```
>> plot(f(1:200), abs(vf(1:200)))
>>
```

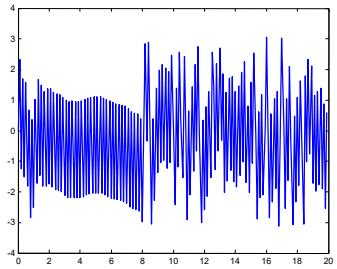
magnitude of the spectrum



MHz

```
>> plot(f(1:200), angle(vf(1:200)))
```

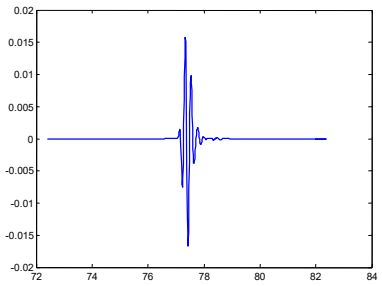
phase of the spectrum (in rad)



MHz

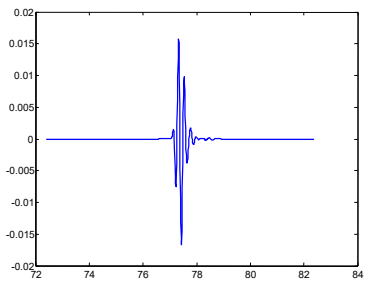
Here we see both the magnitude and phase of the frequency spectrum on a reduced frequency scale. Below 1 MHz and above 8 MHz the spectrum is small so that likely the phase values are contaminated by noise at those frequencies. We see that even in the 1-8 MHz range the phase is rapidly changing. We will examine the reason for this shortly.

```
>> vt =IFourierT(vf, dt);                    inverse FFT
>> plot(texp, vt)
Warning: Imaginary parts of complex X and/or Y arguments
ignored
>> plot(texp, real(vt))
```



µsec

original signal

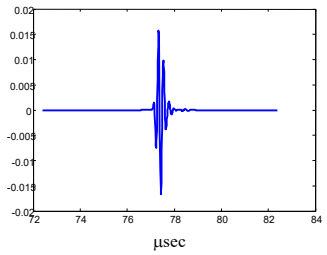


µsec

obtained from inverse FFT

Here we attempt to perform the inverse FFT with the MATLAB function IFourierT to recover our original time domain signal vt. We see that MATLAB gives us a warning that the signal has imaginary parts. These arise because there are always very small round-off errors in doing the complex arithmetic with these transforms, leaving some inconsequential imaginary values. We can remove those values by simply taking the real part of the signal which does recover our original signal.

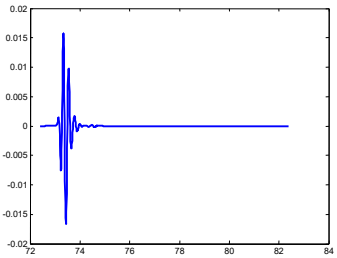
original time signal



µsec

```
>> vshift =[ vexp(400:1000) vexp(1:399)];
>> plot (texp, vshift)
```

shifted time signal

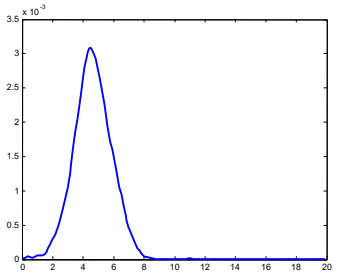


µsec

We saw previously that our signal had a rapidly changing phase. This was due to the fact that the signal was shifted in time to near the center of the plot. If move the signal closer to the left by moving the first 399 points in the beginning of the plot to the end of the plot then we obtain the shifted signal shown.

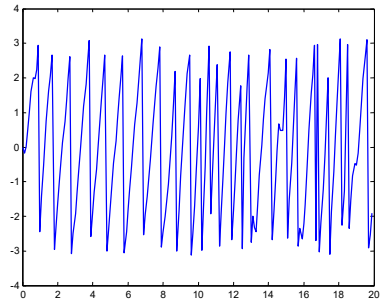
```
>> vfs = FourierT(vshift, dt);
>> plot( f(1:200), abs(vfs(1:200)))
```

same magnitude of spectrum as before



MHz

```
>> plot( f(1:200), angle(vfs(1:200)))
```

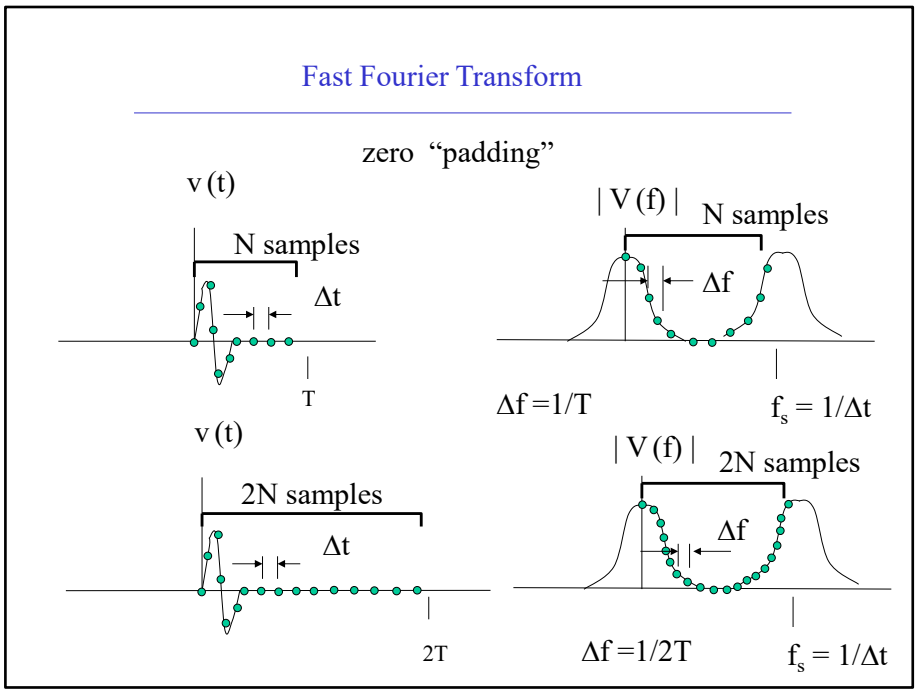


MHz

The phase is now different. This is to be expected because recall a time delay t_0 causes the spectrum to change only in its phase, i.e.

$$v(t) \leftrightarrow V(f)$$
$$v(t-t_0) \leftrightarrow \exp(2\pi i f t_0) V(f)$$

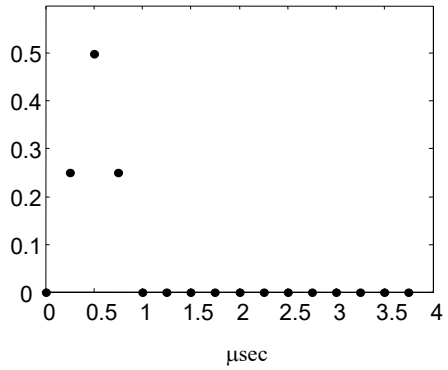
If we now do an FFT of the shifted signal and plot the magnitude and phase of the spectrum over the first 200 points (going from 0 to 20 MHz) we see that the phase is less oscillatory because time delays produce a complex exponential term which affects the phase (but not the magnitude) of the spectrum, as seen above. If we move the signal even further to the left the oscillations will be even less than seen here.



In some cases the sampling time interval (and hence the sampling frequency) may be adequate to prevent aliasing but we would like to increase the number of samples in the frequency domain to better capture any rapidly changing behavior there. We can do this by simply adding (padding) the time domain with zeros spaced at the same sampling time interval. Thus, the sampling frequency is not changed but the number of samples in the frequency domain will be larger. Shown above is an example where we pad a signal with N samples with N additional zeros and hence also double the number of samples in the frequency domain from N to 2N over the same frequency range. This is called **zero padding**.

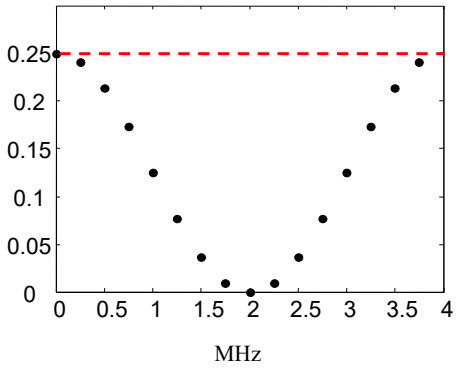
Zero padding example

```
>> t =s_space(0, 4, 16);
>> v = t.*(t < 0.5) + (1-t).*(t >= 0.5 & t<=1.0);
>> h = plot(t, v, 'ko');
>> set(h,  'MarkerFaceColor', 'k')
>> axis([  0  4  0   0.6])

>> dt = t(2) - t(1)

dt = 0.2500
```



As an example, here we generate samples of a triangular time domain function with s-space going from t =0 to t = 4 microseconds. Thus, the sampling interval is 4/16 = ¼ microseconds and the sampling frequency is 4 MHz. Note that s-space properly gives us the correct 16 points and those points do not include a point at t = 4 microseconds as that point is the same as the one at t =0.

```
>> f =s_space(0, 1/dt, 16);
>> vf = FourierT(v, dt);
>> h = plot(f, abs(vf), 'ko');
>>axis([ 0 4 0 0.3])

>> set(h,  'MarkerFaceColor', 'k')
>> df = f(2) - f(1)

df =  0.2500
```

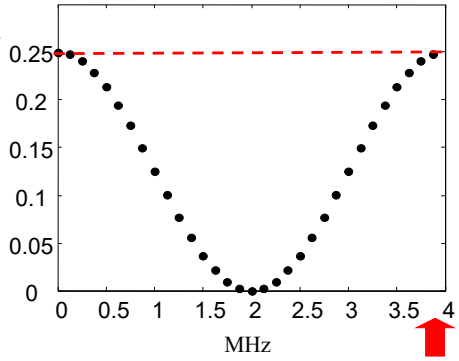


If we compute the FFT of this sampled function and generate a set of sampled frequencies with s_space, then we obtain the 16 sampled frequency domain values as shown with the spacing of ¼ MHz. Note again s_space gives us the proper sampled values and that does not include a value at f = 4 MHz, which is the same as the value at f = 0.

Now, pad the original signal with zeros

```
>> vt2 =[v, zeros(1,16)];
>> vf2 = FourierT(vt2, dt);
>> h = plot(f, abs(vf2), 'ko');
??? Error using ==> plot
Vectors must be the same lengths.

>> f =s_space(0, 1/dt, 32);
>> h = plot(f, abs(vf2), 'ko');
>> set(h,  'MarkerFaceColor', 'k')
>>axis([ 0 4 0 0.3])
>> df=f(2) - f(1)
df = 0.1250
```

could also use
vt2 = padarray(v,[0, 16], 0, 'post');



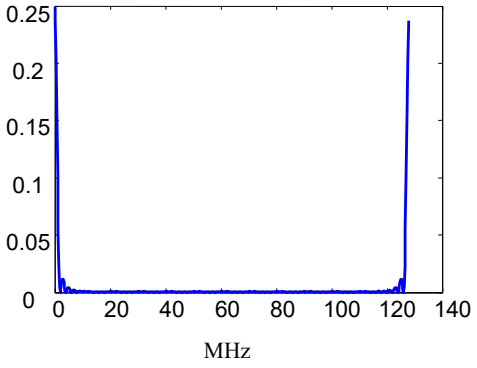
half the former df

same sampling frequency

Now, we pad our original time domain signal by adding 16 zeros and then do a FFT. If we try to plot the result we get an error because our set of frequency values has doubled so we must generate a new set of 32 frequencies from zero to the sampling frequency. Then the plot works fine and we see a much finer resolution where now the frequency domain sampling interval is 1/8 MHz.

Now, use a much higher sampling frequency

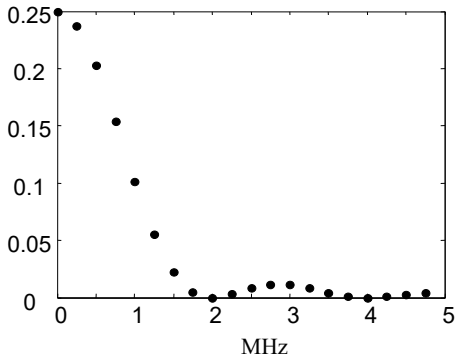
```
>> t =s_space(0, 4, 512);
>> v = t.*(t < 0.5) + (1-t).*(t >= 0.5 & t<=1.0);
>> dt = t(2) -t(1)
dt = 0.0078
>> fs =1/dt
fs = 128

>> f =s_space(0, fs, 512);
>> vf =FourierT(v, dt);
>> plot(f, abs(vf))
```



We used the triangular function example to illustrate zero padding, using very few points so we could easily see the results. However, did we adequately address aliasing? To answer this let us use a much higher sampling frequency (128 MHz = 32 times higher than previously used) as shown here and plot the magnitude of the spectrum. We see in this case the positive and negative frequency values are well separated and the spectral values appear to be small below about 10 MHz. Let us examine those values on a smaller scale (next slide)

To see the frequency spectrum on a finer scale

```
>> h = plot(f(1:20), abs(vf(1:20)), 'ko');
>> set(h,  'MarkerFaceColor', 'k')
```

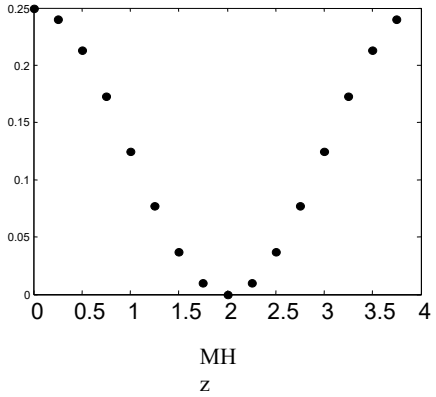


Note: we had some aliasing before

If we plot the first 20 values of the spectrum (which are values from approximately 0 to 5 MHz) we see that there are some small higher frequency values that were missed previously so a sampling frequency of 4 MHz was not sufficient to completely prevent aliasing.

We can do multiple FFTs or IFFTs
all at once if we place the data in columns

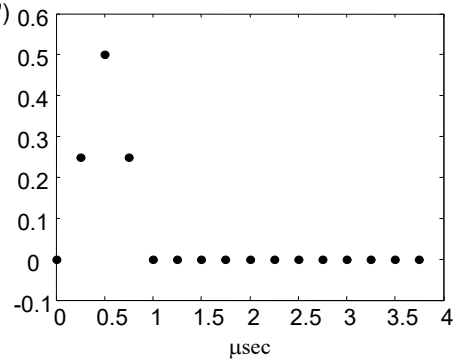
```
>> t =s_space(0, 4, 16);
>> v = t.*(t < 0.5) + (1-t).*(t >= 0.5 & t<=1.0);
>> dt = t(2) - t(1);
>> mv= [ v'  v'  v'];
>> mvf = FourierT(mv, dt);
>> vf1 = mvf(:,1);
>> f = s_space(0, 1/dt, 16);
>> h = plot(f, abs(vf1)', 'ko')
>> set(h, 'MarkerFaceColor', 'k')
```



Note that we can do multiple FFTs or inverse FFTs all at once for multiple signals if we place those sampled signals in a matrix where the sampled values are in the columns of that matrix. Here we generate such a matrix which simply contains the original triangular function used previously but duplicated three times in the columns of the matrix mv. We then do the FFT of that matrix and extract the first column of frequency domain values in the matrix mvf and plot them. Identical values are obviously contained in the other two columns.

Note that even though there is aliasing here if we do the inverse FFT of these samples we do recover the original time samples:

```
>> v1 =IFourierT(vf1,dt);
>> h=plot(t, real(v1), 'ko');
>> set(h, 'MarkerFaceColor', 'k')
```



Recall that the sampling used in the previous slide was insufficient to prevent aliasing. In spite of that fact, when we perform an inverse FFT we do obtain the proper time domain function. Thus, aliasing generates incorrect frequency domain values through the FFT but does not prevent us from recovering the original function with the inverse FFT.

Fast Fourier Transform

FFT Examples using the function:

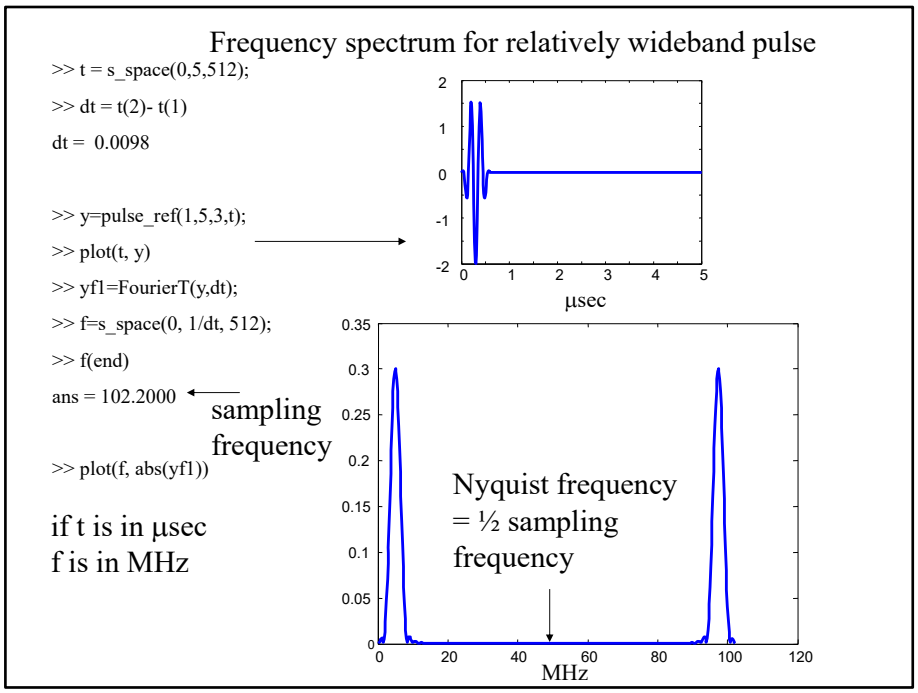
$$y(t) = \begin{cases} A\left[1-\cos\left(2\pi Ft/N\right)\right]\cos\left(2\pi Ft\right) & \left(0 < t < N/F\right) \\ 0 \quad otherwise \end{cases}$$

A … controls the amplitude
F … controls the dominant frequency in the pulse
N … controls the number of cycles (amount of "ringing") in the pulse and hence the bandwidth

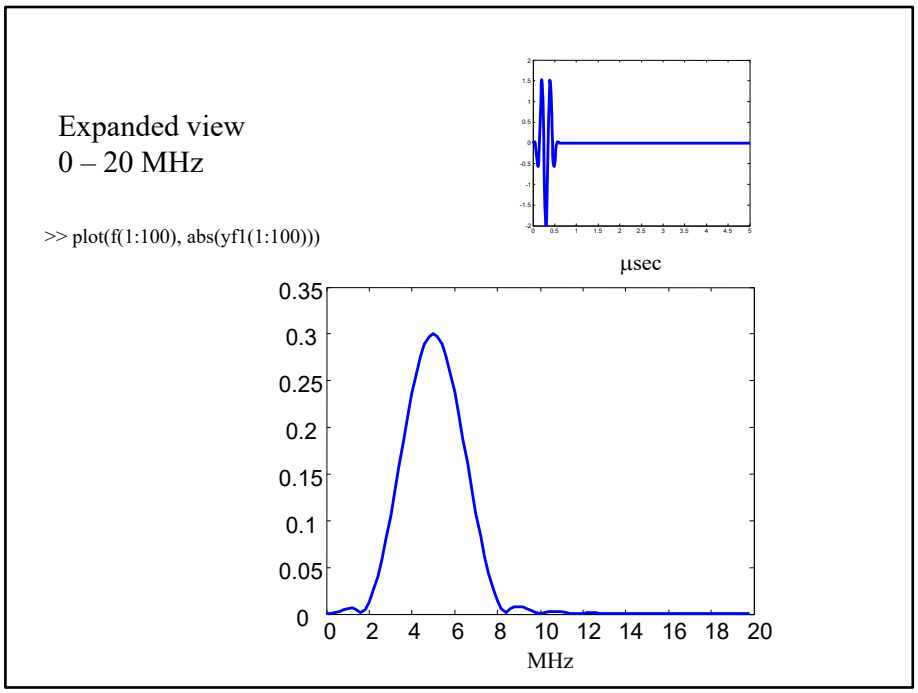
MATLAB function:

```
function y = pulse_ref(A,F,N, t)
y = A*(1 -cos(2*pi*F*t./N)).*cos(2*pi*F*t).*(t >= 0 & t <= N/F);
```

Here, we will define a function that is controlled by three parameters (A, F, N) that will allow us to demonstrate some explicit results.

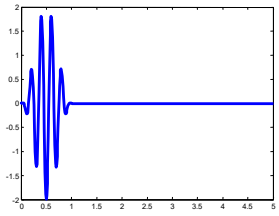


If we let A =1 , f = 5 , N =3 we will generate a short signal with little ringing that has a relatively broad spectrum. As shown we will choose a sampling frequency of about 100 MHz which from the spectrum seen is vey conservative.



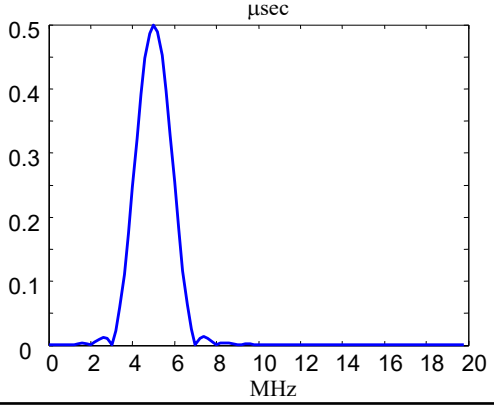
If we look at the spectrum on a finer scale we see a relatively broad spectrum centered at 5 MHz.

```
>> y = pulse_ref(1,5,5,t);
>> yf2 = FourierT(y, dt);
>> plot(f(1:100), abs(yf2(1:100)))
```

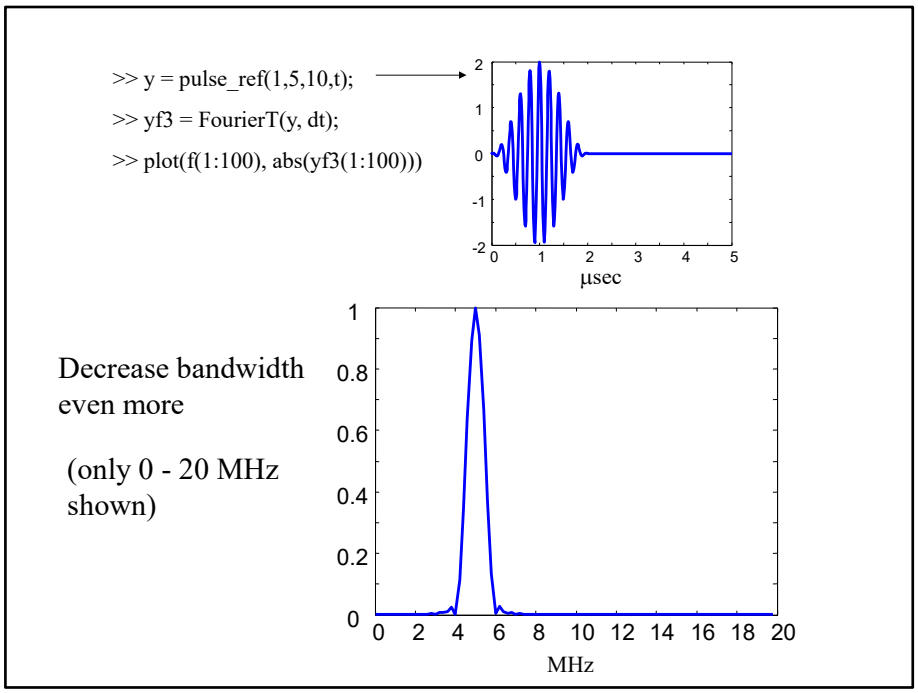


μsec

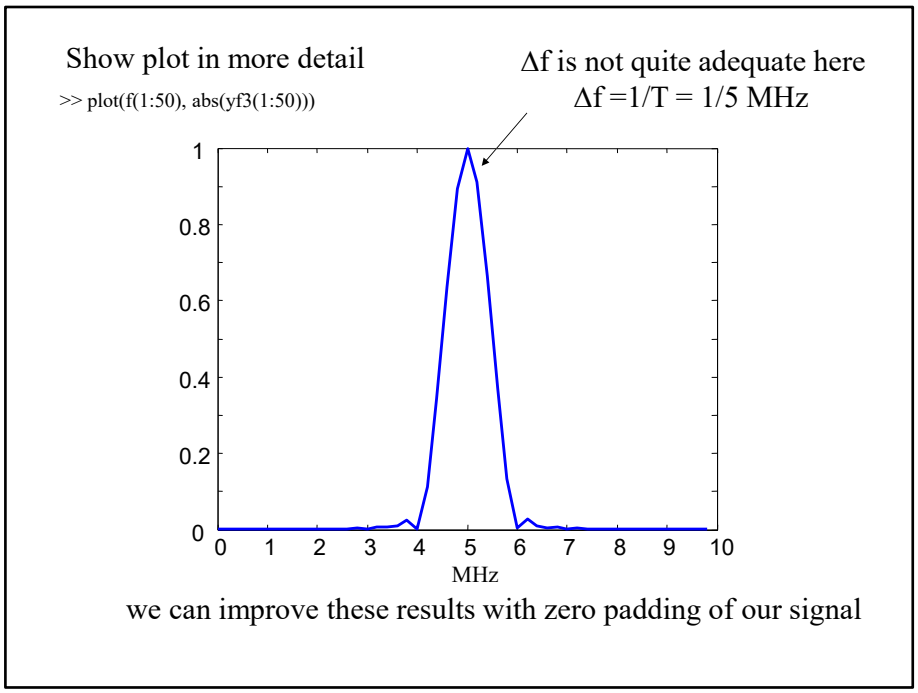
Somewhat more narrow band pulse (only 0 - 20 MHz shown)



If we increase the amount of “ringing” in the signal by making N larger (N =5) we see a somewhat narrower bandwidth for the spectrum.



If we increase the amount of ringing even more (N = 10) we see an even narrower spectrum



If we examine the spectrum of this narrowband signal on a finer scale we see that the spectrum is jagged, indicating that our spacing in the frequency domain is not adequate. We can use zero padding to improve the results, as we will now show.

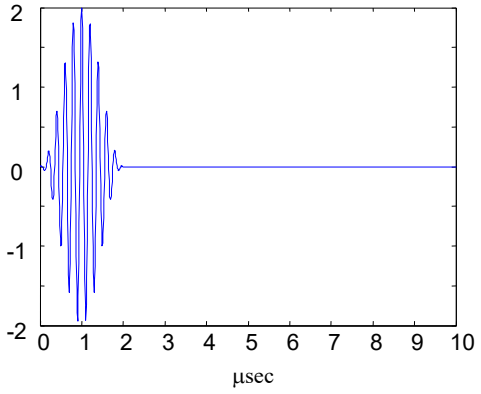
zero padding

```
>> t=s_space(0, 10, 1024);
>> y=pulse_ref(1, 5,10, t);
>> plot(t,y)
>> dt = t(2) -t(1)

dt = 0.0098
```

recall, originally, we had

t = s_space(0,5, 512);

we could also zero pad via
t = [t, zeros(1, 512)];

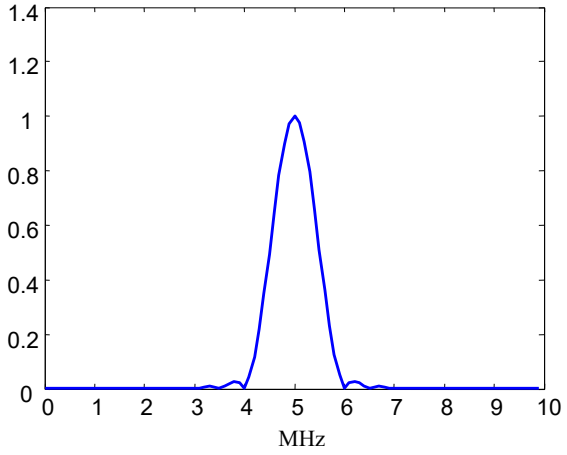


now, have 1024 points over 10 microseconds, so dt is same but T = Ndt is twice as large

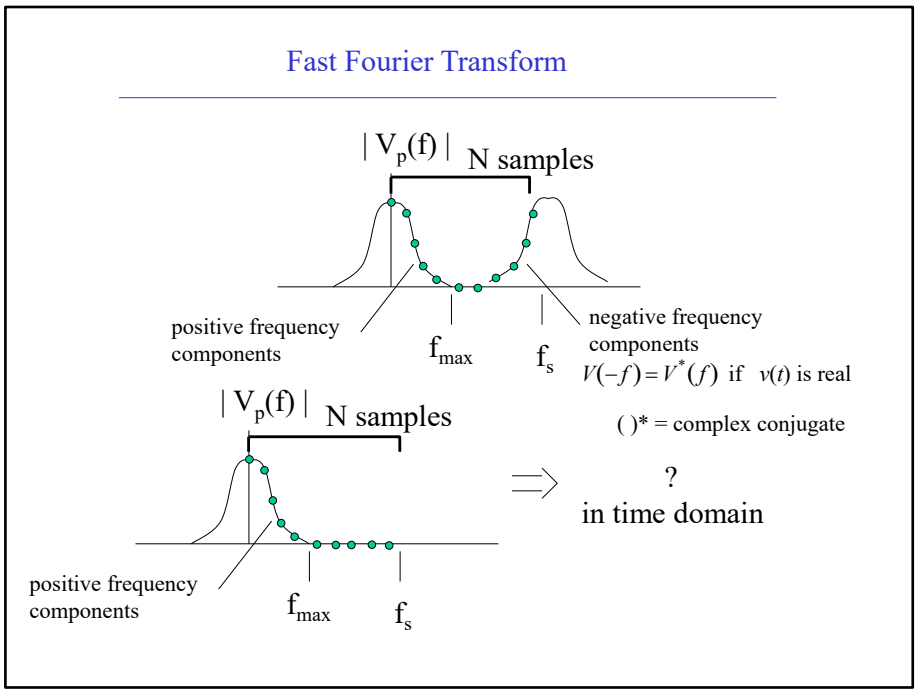
Here we double the number of points with zero padding but keep the same sampling frequency.

```
>> yf4 = FourierT(y, dt);
>> f = s_space(0, 1/dt, 1024);
>> plot(f(1:100), abs(yf4(1:100)))
```

Improved representation of the spectrum $\Delta f = 1/T = 1/10$ MHz



Now if we examine the spectrum we see a smoother signal. We can improve the results even more by adding more zeros.



When we perform an FFT we see that in the frequency domain we obtain the positive frequency components in the lower half of the spectrum and the negative frequency components in the upper half of the spectrum. The negative frequency values are just the complex conjugate of the positive frequency components so they do not contain any new information. They exist solely to guarantee that the original time domain signal was a real signal. **What happens if we simply replace the negative frequency values by zeros and then do an inverse FFT?**

Fast Fourier Transform

$$v(t) = \int_{-\infty}^{+\infty} V(f)\exp(-2\pi i f t)df$$

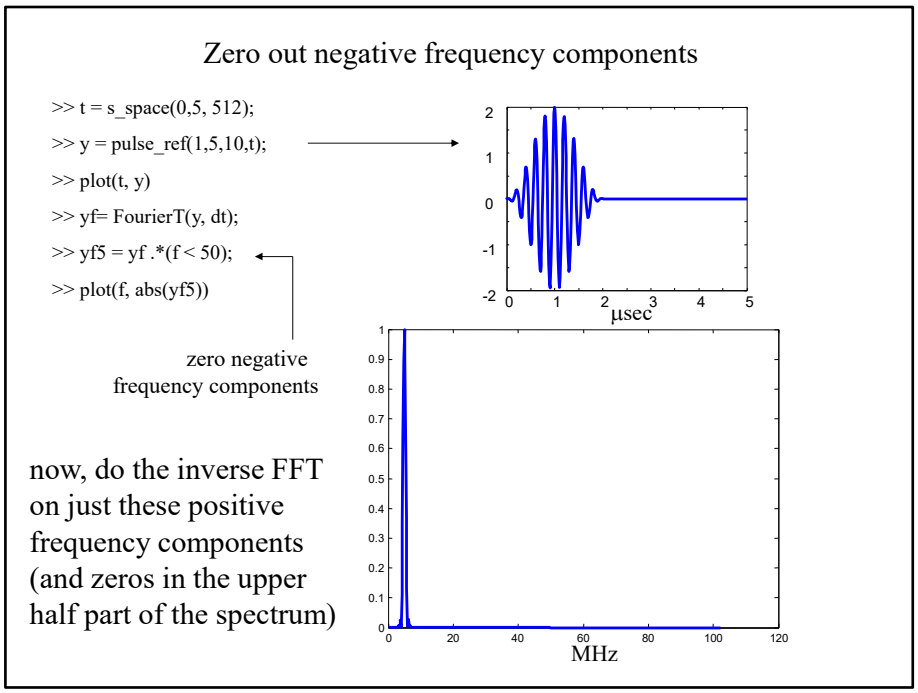
$$\frac{v(t)}{2} - \frac{i}{2}H[v(t)] = \int_{0}^{+\infty} V(f)\exp(-2\pi i f t)df$$

Hilbert transform of v(t)

so

$$v(t) = 2\,\mathrm{Re}\left\{\int_{0}^{+\infty} V(f)\exp(-2\pi i f t)df\right\}$$

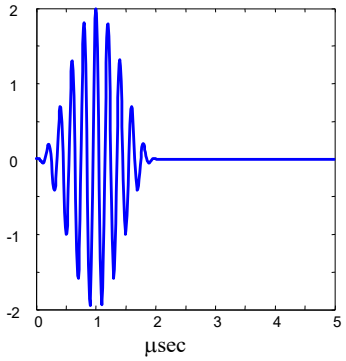
The answer is we obtain a complex signal where one half of our original time domain signal is in the real part and the imaginary part contains minus one half of the Hilbert transform of our original time domain signal. Thus, we can replace the upper half our spectrum by zeros and recover our original real time domain signal by just taking twice the real part of the inverse FFT of the remaining positive frequency components.



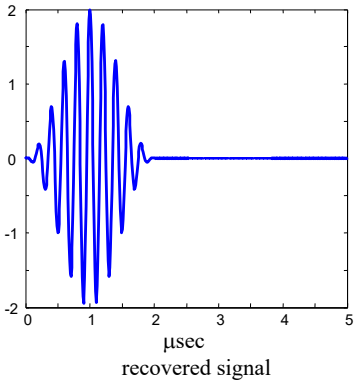
Here is an example where we compute the spectrum of signal and replace the negative frequency values by zeros. However, what about the zero frequency value in the spectrum? Do we consider it a positive or negative value? The answer is: we consider half of it positive and half of it negative so that we divide the zero frequency value of the spectrum (if it is non-zero) by two before we do an inverse FFT.

```
>> yf5(1) = yf5(1)/2;
>> y = 2*real(IFourierT(yf5, dt));
>> plot(t, y)
```

when throwing out negative frequency components, need to divide dc value by half also (not needed if dc value is zero or already very small but it is a good idea to always do it)

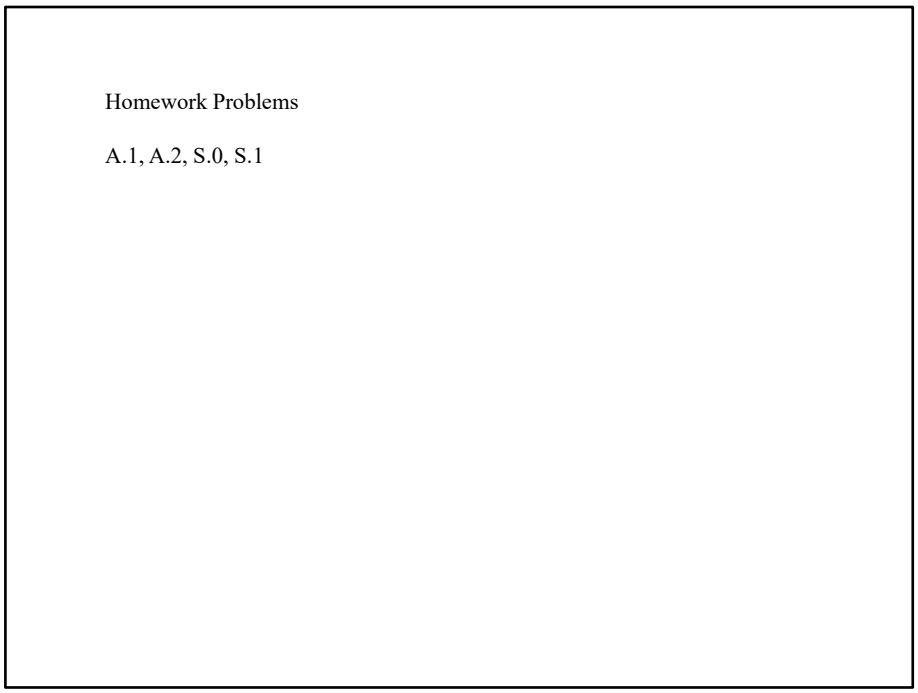


original signal



recovered signal
from real frequency components

Here we see the zero frequency value being divided by two and then twice the real part of the inverse FFT is calculated. We see we do recover the original signal



Special homework problems and problems from Appendix A.

Fast Fourier Transform

References

Walker, J.S., **Fast Fourier Transforms**, CRC Press, 1996

Burrus, C.S. and T.W. Parks, **DFT/FFT and Convolution Algorithms**, John Wiley and Sons, New York, 1985.

There are many references on Fourier transforms and the fast Fourier transform. Here are several of them.

Impedance Concepts and Thévenin Equivalent Systems

We will see several examples where the concept of impedance is important and examine the concept of Thevenin equivalent systems.

Learning Objectives

Electrical impedance

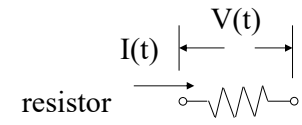
Description of 1-D waves in a fluid
- equation of motion/constitutive equation
- plane waves - pulses and harmonic waves
- frequency and wavelength
- acoustic impedance

Thévenin equivalent electrical circuits

Electrical impedance is a property you may already be familiar with. We will see that impedance is also defined for the waves present in ultrasonic system and examine some basic characteristics of waves. Finally, we will discuss the concept of Thevenin equivalent electrical circuits.

Concept of Impedance

basic circuit elements



resistor $V(t) = R\, I(t)$

capacitor $C\dfrac{dV(t)}{dt} = I(t)$

inductor $V(t) = L\dfrac{dI(t)}{dt}$

Some of the basic elements often seen in circuits include resistors, capacitors, and inductors. In each of these elements we can define the relationship between the voltage acting across the element in terms of the current flowing through the element.

For alternating (harmonic) voltages and currents

$I(t) = I_0 \exp(-i\omega t)$

$V(t) = V_0 \exp(-i\omega t)$

resistor $\quad V_0 = R\, I_0$

capacitor $\quad V_0 = \dfrac{I_0}{-i\omega C}$

inductor $\quad V_0 = -i\omega L\, I_0$

general impedance $\quad V_0 = Z(\omega)\, I_0$

electrical impedance Z = Z^e (ohms)

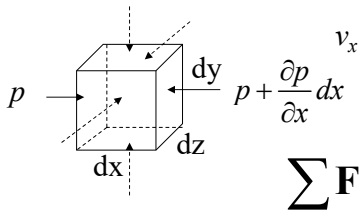
For alternating voltages and currents we can relate the voltages and current in these elements through a complex electrical impedance, Z, which in general can also be a function of the frequency. We will place an e superscript on Z to indicate it an **electrical impedance** and to distinguish it from other impedances we will discuss shortly.

1-D plane wave traveling wave in a fluid

$p(x, t)$ … pressure in the fluid

ρ … fluid density

$v_x(x, t)$ … velocity



$$\sum \mathbf{F} = m\mathbf{a}$$

$$-\left(p+\frac{\partial p}{\partial x}dx\right)dydz + pdydz = \rho dxdydz\frac{\partial v_x}{\partial t}$$

Equation of motion $$-\frac{\partial p}{\partial x} = \rho\frac{\partial v_x}{\partial t}$$

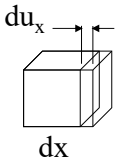
We will see that impedance is also defined for other types of non-electrical problems. Since ultrasonic NDE involves waves propagating in fluids and solids, let us consider some aspects of wave propagation before we define the concept of wave impedance. First, let us consider the problem of waves propagating in a fluid where all the motion of the wave is in the x-direction and the pressure in the wave only depends on x . Since the pressure is a constant in the y-z plane this is a 1-D plane wave. If we examine Newtons law F = ma for a small element of the fluid we can derive the equation of motion which relates the derivative of the pressure to the acceleration of the fluid element.

Constitutive equation

$$p = -K\frac{\partial u_x}{\partial x}$$

K … compressibility of the fluid

u_x … displacement



$$\frac{\partial u_x}{\partial x} = \frac{\Delta V}{V}$$ = relative change in volume (volumetric strain)

$$-\frac{\partial^2 p}{\partial x^2} = \rho\frac{\partial^2\left(\partial u_x/\partial x\right)}{\partial t^2}$$

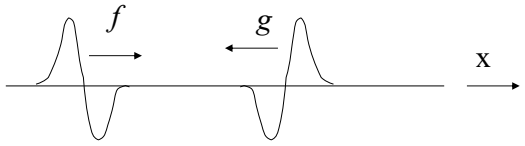
$$= \frac{\rho}{K}\frac{\partial^2 p}{\partial t^2}$$

1-D wave equation $\frac{\partial^2 p}{\partial x^2} - \frac{1}{c_f^2}\frac{\partial^2 p}{\partial t^2} = 0$ $c_f = \sqrt{\frac{K}{\rho}}$ wave speed in the fluid

For an ideal fluid the pressure is linearly related to the displacement gradient of the element which is also equal to the so-called volumetric strain. The proportionality constant in this constitutive equation is called the **compressibility**, K, of the element. If we take the derivative of Newtons law with respect to x and use the constitutive equation we arrive at the 1-D wave equation for the pressure, where a new constant, called the wave speed , appears.

1-D plane wave solutions

$$p = f\left(t - x/c_f\right) + g(t + x/c_f)$$



Harmonic waves

$$p = A(f)\exp\left[-2\pi i f\left(t - x/c_f\right)\right] + B(f)\exp\left[-2\pi i f\left(t + x/c_f\right)\right]$$

The 1-D wave equation has 1-D traveling plane wave solutions where we can have pressure waves defined by arbitrary functions that travel in the plus or minus x-direction. As a special case we can let those functions be harmonic traveling waves.

Fourier transform pair

$$V(f) = \int_{-\infty}^{+\infty} v(t) \exp(2\pi i f t) dt$$

$$v(t) = \int_{-\infty}^{+\infty} V(f) \exp(-2\pi i f t) df$$

Let $V(f) = \tilde{V}(f) \exp\left[2\pi i f \left(x / c_f\right)\right]$ (1)

then $v(t) = \int_{-\infty}^{+\infty} \tilde{V}(f) \exp\left[2\pi i f \left(x / c_f - t\right)\right] df$

1-D plane harmonic wave traveling in the +x-direction

Thus, *v(t)* is a superposition of plane waves

Now let us examine the Fourier transform and inverse Fourier transforms for a spectrum that has the particular harmonic wave form shown in Eq. (1). Then we see the inverse Fourier transform represents a superposition of 1-D traveling harmonic waves. What is the corresponding time domain signal that we recover from this inverse Fourier transform?

Now, let $u = \left(t - x/c_f\right)$

Then $\int_{-\infty}^{+\infty} \tilde{V}(f)\exp[-2\pi i f u]\,df = \tilde{v}(u) = \tilde{v}\left(t - x/c_f\right)$

where $\tilde{v}(t) \leftrightarrow \tilde{V}(f)$

so

$$\tilde{v}\left(t - x/c_f\right) = \int_{-\infty}^{+\infty} \tilde{V}(f)\exp\left[2\pi i f\left(x/c_f - t\right)\right]df$$

Thus, we can always synthesize a traveling plane wave pulse with a superposition of harmonic plane waves.

By simply changing to a different variable we can see that if a time domain waveform v(t) has a frequency domain spectrum V(f) then if we superimpose traveling harmonic waves with amplitude V(f) we recover from the inverse FFT a traveling pulse with the same waveform as v. This is shown above for a wave traveling in the plus x-direction but a similar result holds for a wave traveling in the negative x-direction. Thus, we can always work with harmonic waves in describing NDE systems since we can recover the corresponding pulses present in those systems simply by performing an inverse FFT of the harmonic wave solutions.

various forms of a harmonic plane wave traveling in the +x-direction

$$p = A\exp\left(2\pi ifx/c_f - 2\pi ift\right)$$
$$= A\exp\left(2\pi ifx/c_f - i\omega t\right) \qquad f = \omega/2\pi$$
$$= A\exp\left(ikx - i\omega t\right) \qquad k = \omega/c_f$$
$$= A\exp\left(2\pi ix/\lambda - i\omega t\right) \qquad \lambda = 2\pi/k = c_f/f$$

f ... frequency (cycles/sec)

ω …frequency (rad/sec)

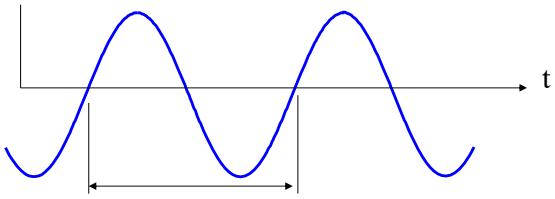
k … wave number (rad/length)

λ …wavelength (length/cycle)

A traveling harmonic plane wave can be written in a number of different forms, all of which are equivalent. The form indicated by the red arrow is one form uwe will use frequently in solving wave problems.

$$p = A\exp\left(2\pi ifx/c_f - 2\pi ift\right)$$

magnitude of the pressure at a fixed x versus time, t:



T… period (sec/cycle)

$$2\pi ifT = 2\pi i \quad (1)$$

$$f = \frac{1}{T}$$ frequency (cycles/sec)

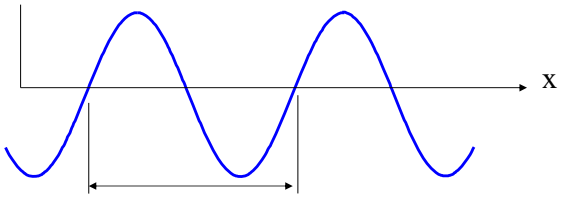
$$\omega = 2\pi f$$ circular frequency (rad/sec)

rad/cycle

The period T of a harmonic wave is the time it takes for the waveform to go through a complete cycle of 2π radians so by examining the phase term involving time in the wave (see Eq. (1)) as we hold x fixed we obtain the result that the frequency (in cycles/sec) is just the reciprocal of the period. The circular frequency is just 2π times the frequency in cycles/sec since there are 2π radians in a cycle.

$$p = A\exp\left(2\pi ifx/c_f - 2\pi ift\right)$$

magnitude of the pressure at a fixed time, t, versus distance, x:



λ ... wavelength (length/cycle)

$f_x = \frac{1}{\lambda}$ spatial frequency (cycles/length)

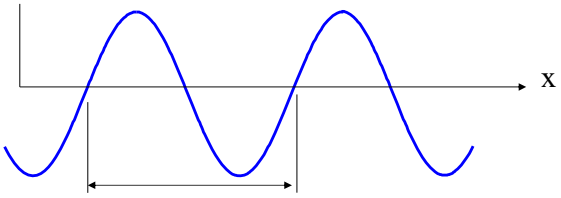
$k = 2\pi f_x = \frac{2\pi}{\lambda}$ spatial circular frequency (wave number) (rad/length)

↑ rad/cycle

If instead we hold the time fixed and plot the pressure versus x we again seen a sinusoidal curve. The distance during one cycle is the **wavelength** which acts like the period does in time so the reciprocal of the wavelength acts like a spatial frequency. If we multiply that spatial frequency by 2π we get a spatial circular frequency called the **wavenumber**, k.

$$p = A\exp\left(2\pi ifx/c_f - 2\pi ift\right)$$

magnitude of the pressure at a fixed time, t, versus distance, x:



λ ... wavelength (length/cycle)

$$2\pi i f\lambda/c_f = 2\pi i$$

$$\Rightarrow \quad f\lambda = c_f \qquad (1)$$

fundamental relationship that shows how the frequency of a plane harmonic traveling wave is related to its wavelength

If we hold the time fixed and vary the x-location over one complete cycle (2pi radians), by definition we go through a distance, λ , equal to the wavelength of the wave, so that from Eq. (1) we find that the frequency (in cycles/sec) time the wavelength is just the wave speed.

Example:

consider a 5MHz plane wave traveling in water (c =1500m/sec). What is the wave length?

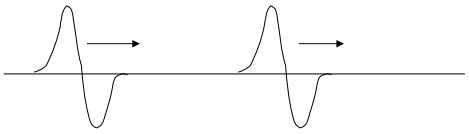
$$\lambda = \frac{1.5 \times 10^6 \, mm/\sec}{5 \times 10^6 \, cycle/\sec}$$
$$= 0.3 mm/cycle$$

Here is an example of the wavelength for a 5 MHz wave traveling in water. Thus, we see the wavelengths in NDE tests are typically quite small.

For a plane pressure wave traveling in the x-direction

$$p = f\left(t - x/c_f\right)$$

Let $u = t - x/c_f$

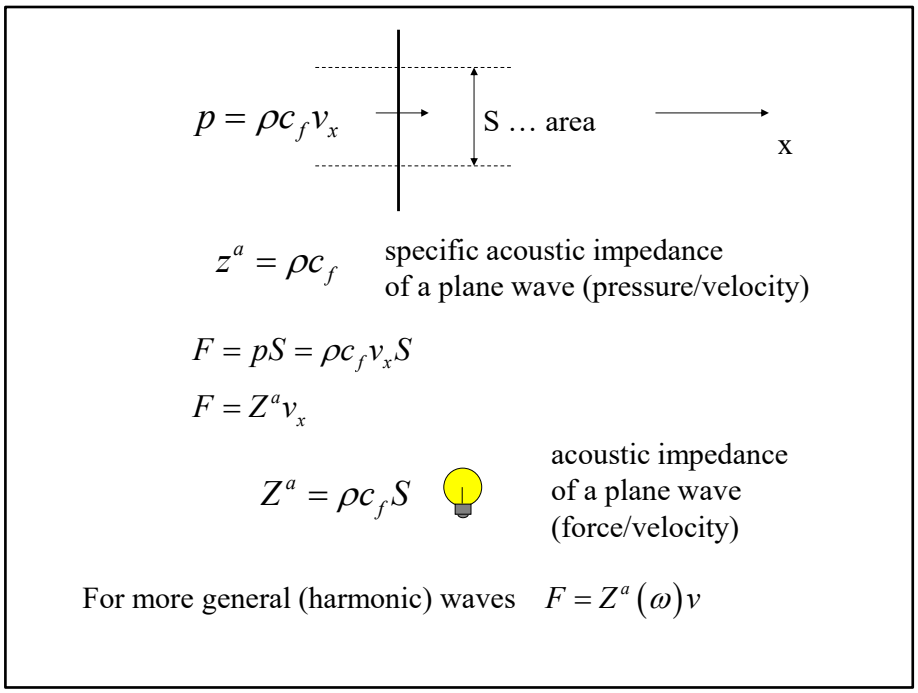


then $\dfrac{\partial p}{\partial x} = \dfrac{df(u)}{du}\dfrac{\partial u}{\partial x} = -\dfrac{1}{c_f} f'$

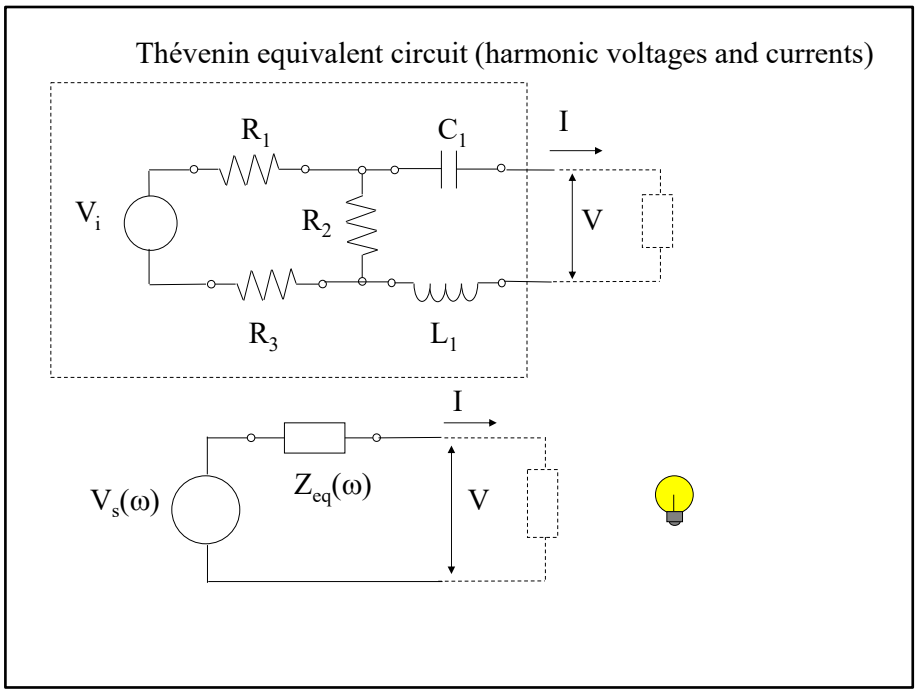
$$-\frac{\partial p}{\partial x} = \rho \frac{\partial v_x}{\partial t} \quad \Rightarrow \quad \frac{\partial v_x}{\partial t} = \frac{1}{\rho c_f} f'$$

$$v_x = \frac{1}{\rho c_f}\int f' \partial t = \frac{1}{\rho c_f}\int \frac{df}{du} du = \frac{f}{\rho c_f} = \frac{p}{\rho c_f}$$

If we consider a plane wave traveling in the x-direction, we can place that plane wave into Newton's law and integrate to find the x-velocity, which we see is just proportional to the pressure

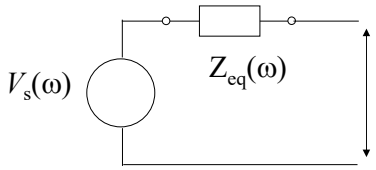


The ratio of the pressure to the velocity is a quantity called the **specific acoustic impedance** of the plane wave. If we multiply that specific acoustic impedance by an area of the wavefront then the ratio of the force in the wave over that area to the velocity is called the **acoustic impedance** of the plane wave. This impedance, F/v, plays the same role for the wave as the voltage to current ratio, V/I, plays in an electrical circuit. To distinguish this impedance, Z, from an electrical impedance we will place an "a" superscript on it to indicate it is an acoustic impedance. Although the acoustic impedance of a plane wave is a constant, for other types of waves the acoustic impedance, like the electrical impedance, can be a function of the frequency.



The reason that the concept of impedance is so important is because if we take a linear electrical system of components, such as the one shown, we can replace it by an equivalent system consisting only of a voltage source and an impedance, called a **Thevenin equivalent system**. Both systems are equivalent because if both systems are terminated in some fashion (as shown by the dotted lines) the voltage and current outputs are identical. We will not prove this equivalence here.

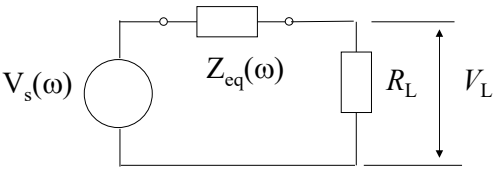
Determining the Thévenin equivalent source



open-circuit voltage

$$V_s = V_0$$

Determining the Thévenin equivalent impedance



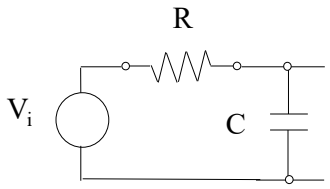
$$V_s - V_L = Z_{eq}^e I$$

$$V_L = R_L I$$

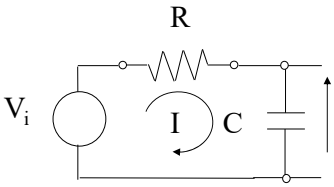
$$(1)\ Z_{eq} = R_L\left(\frac{V_s}{V_L} - 1\right)$$

We can determine the Thevenin equivalent system in two steps. First, if we measure the open circuit output voltage of the system, this will just be the equivalent voltage source acting. Then if we place a known impedance (such as the resistance shown above) at the output and measure the output voltage across this known resistance, then we can find the equivalent impedance as shown in Eq. (1) above.

Example: find the Thévenin equivalent circuit for the following circuit



As an example of determining a Thevenin equivalent system, consider a simple circuit consisting of a known voltage source and a resistance and capacitance.



V_0 open-circuit voltage

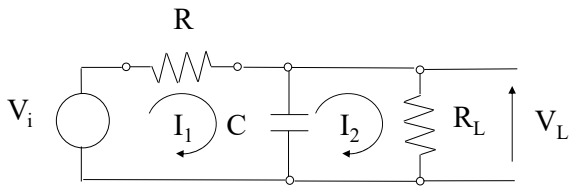
$$V_i - V_0 = IR$$

$$V_0 = \frac{I}{-i\omega C}$$

eliminating I, we find

$$V_0 = \frac{V_i}{1 - i\omega RC}$$ Thévenin equivalent source

If we examine the voltages and current acting across the resistor and capacitor under open circuit conditions then we can eliminate the current and obtain an expression for the open circuit voltage, which is just the Thevenin equivalent voltage source.



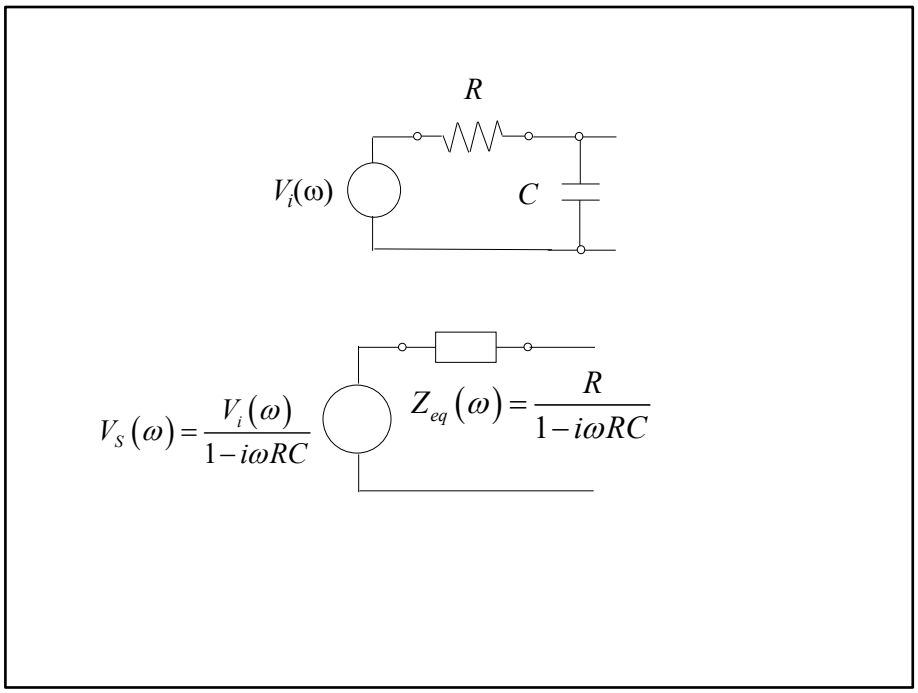
$$V_i - V_L = I_1 R \qquad V_L = I_2 R_L \qquad V_L = \frac{(I_1 - I_2)}{-i\omega C}$$

eliminating I_1 , I_2 gives $\quad V_L = \dfrac{V_i}{(1 - i\omega RC) + R/R_L}$

so (1) $\quad Z_{eq} = R_L \left(\dfrac{V_0}{V_L} - 1 \right) = R_L \left\{ \dfrac{(1 - i\omega RC) + R/R_L}{(1 - i\omega RC)} - 1 \right\}$

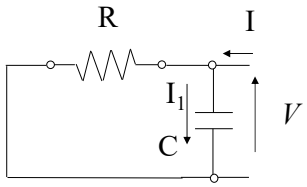
$$= R_L \left\{ \frac{R/R_L}{(1 - i\omega RC)} \right\} = \frac{R}{(1 - i\omega RC)}$$

Similarly, if we place a known resistance at the output terminals and relate the voltages across the resistances and capacitance to the currents flowing, we can eliminate those currents and find the voltage ,V_L, across the known resistance. Equation (1) then gives us the Thevenin equivalent impedance



Thus, the original electrical system is equivalent to the Thevenin equivalent system shown.

In many EE books the equivalent impedance is obtained by simply shorting out (removing) the source and examining the ratio V/I at the output port:



$$V = R\left(I - I_1\right)$$

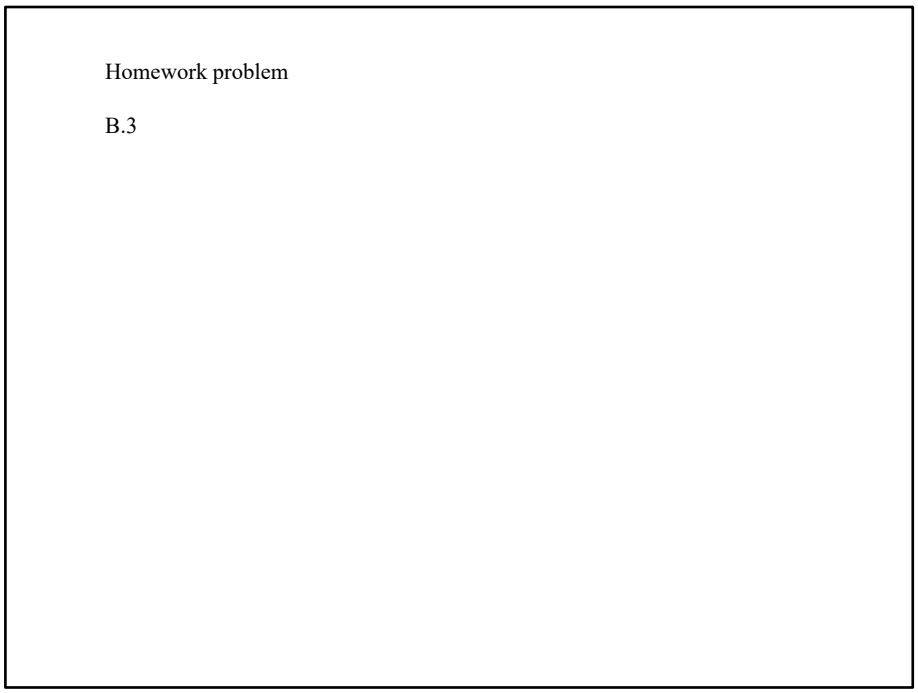
$$V = \frac{I_1}{-i\omega C}$$

$$V = R\left(I + i\omega C V\right)$$

$$V\left(1 - i\omega RC\right) = RI$$

$$Z_{eq} = \frac{V}{I} = \frac{R}{1 - i\omega RC}$$

In an electrical engineering course on circuits you may see another approach to obtaining the Thevenin equivalent impedance where we imagine shorting out the known source and then simply examine the ratio of the output voltage and current. As shown above, this does give the correct impedance but it obviously is not practical (or wise) to physically short out the source in a real instrument. Experimentally, we must use the previous approach, which only requires the measurement of V_L and the open circuit voltage.



Homework problem from Appendix B.

Pulser Characteristics
- Models and Measurements

This section will examine the pulser section of the pulser/receiver

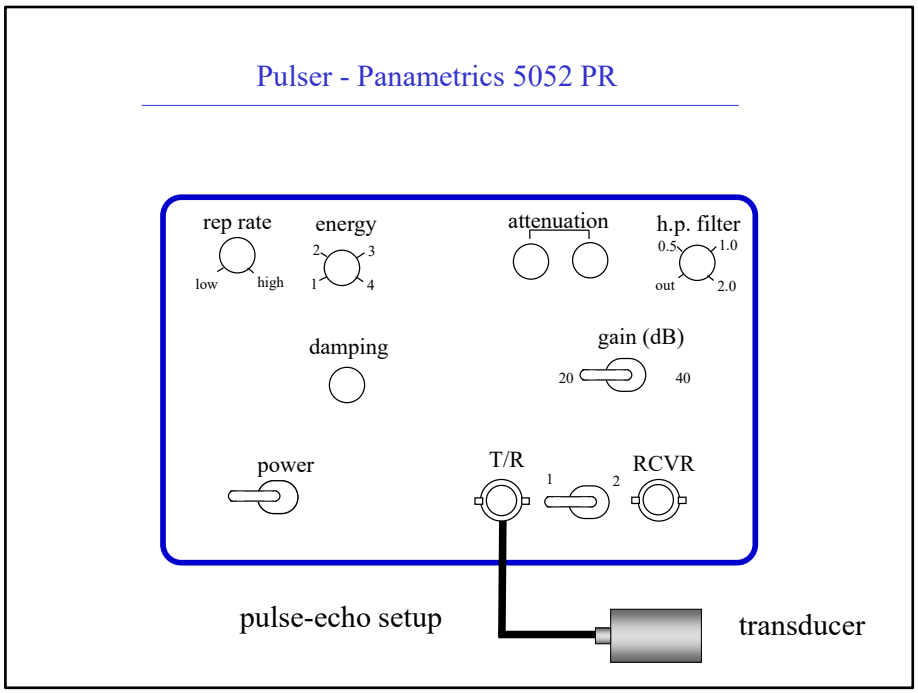
Learning Objectives

characteristics of "spike" and square wave pulsers

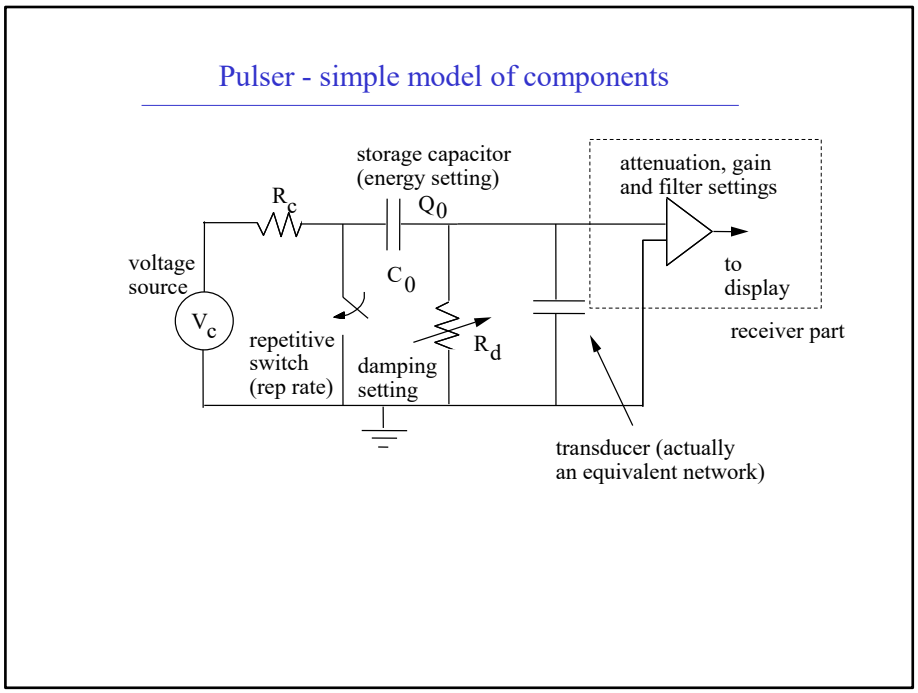
measurement of pulser electrical properties

effects of pulser settings

We will examine two types of pulsers used in practice – a “spike” pulser and a square wave pulser and discuss the measurement of the electrical characteristics of both of these types of instruments. We will also show some examples of the effects of different pulser settings.

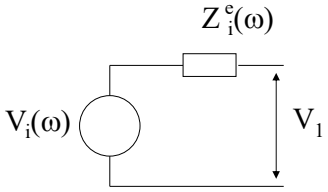


Here is a diagram of a very simple pulser/receiver used in lab settings shown with an attached transducer which is being used as both a sender and receiver of ultrasound. The pulser settings on this instrument consist of the repetition rate setting, an energy level setting, and a damping control.



Here is a very basic model of what is in the pulser/receiver. A voltage source charges up a capacitor which then is periodically discharged. The amount of capacitance, and hence the energy stored, is controlled by the **energy** setting. The **rep rate** controls this discharge rate by the rate at which a switch is closed. One may want to reduce the rep rate in cases where one is inspecting a very large range of depths in a component so that all the flaw or other responses being recorded have attenuated and so that there is no overlap between the driving pulses and the received signal pulses. The **damping** setting controls a variable resistor where a low setting is for a high resistance and vice-versa. Although the transducer is a complex electromechanical device, because it consists of a piezoelectric crystal which is plated on its faces, it acts to first order like a capacitor, which is how it is characterized here. Obviously, the actual pulser circuits are much more complex than those shown here.

Pulser – Thevenin equivalent circuit



$V_i(\omega)$ … equivalent voltage source

$Z^e_i(\omega)$ … equivalent impedance

Even though the pulser circuits may be very complex, if we assume the pulser acts as a linear system then we can replace it by a simpler Thevenin equivalent circuit consisting of a voltage source and an impedance.

Pulser Modeling

Many modeling studies omit the impedance and take the voltage source as a negative time-domain "spike":

Four parameter model:

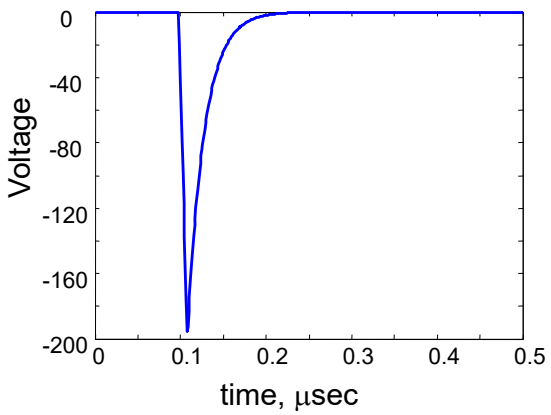
$(t_0, \alpha_1, \alpha_2, V_0)$

$$V_i(t) = \begin{cases} 0 & t \le 0 \\ -V_\infty \left[1 - \exp(-\alpha_1 t)\right] & 0 \le t \le t_0 \\ -V_0 \exp\left[-\alpha_2 (t - t_0)\right] & t \ge t_0 \end{cases}$$

$$V_\infty = \frac{V_0}{\left(1 - e^{-\alpha_1 t_0}\right)}$$

A pulser such as the Panametrics 5052PR puts out very short "spikes" of voltage so it is called a **spike pulser**. Some modeling studies try to simulate the pulser output with a simple model such as the four parameter model shown here.

Pulser Modeling

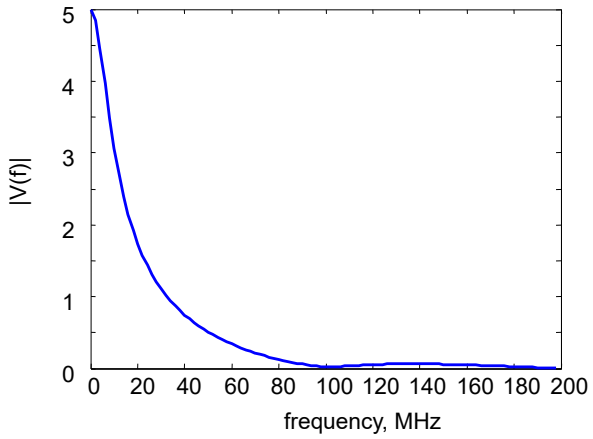


Model pulse with $t_0 = 0.01$, $\alpha_1 = 0.2$, $\alpha_2 = 50$, $V_0 = 200$

Here is an example of the voltage pulse generated with this four parameter model for a set of parameters that mimic roughly what one sees as the output of the pulser: a downward going spike of about .05 microseconds duration and about 200 volts in amplitude.

Pulser Modeling

Frequency spectrum calculated with FFT looks like:



If one computes the frequency components of this pulse with an FFT, here is what the magnitude of the frequency components look like.

```
>> t = linspace(0, 0.5, 512);
>> V = pulserVT(200, .005, 0.2, 50,t);
>> plot(t, V)
>> plot(t, c_shift(V, 100))
>> V = pulserVT(200, .01, 0.2, 50,t);
>> plot(t, c_shift(V, 100))
>> xlabel(' Time (\musec)')
>> ylabel('Voltage')
>> dt = t(2) -t(1);
>> Vf=FourierT(V,dt);
>> f = linspace(0, 1/dt, 512);
>> plot(f, abs(Vf))
>> plot(f(1:100), abs(Vf(1:100)))
>> xlabel('frequency, Hz')
>> ylabel('|V(f)|')
```

```
function V = pulserVT(V0, t0, a1, a2, t)
t = t + eps*( t ==0);
Vinf = V0/(1-exp(-a1*t0));
V = -Vinf*(1- exp(-a1*t)).*(t <= t0) -V0*exp(-a2*(t -t0)).*(t > t0);
```

Here is the MATLAB code that generates these time domain and frequency domain results.

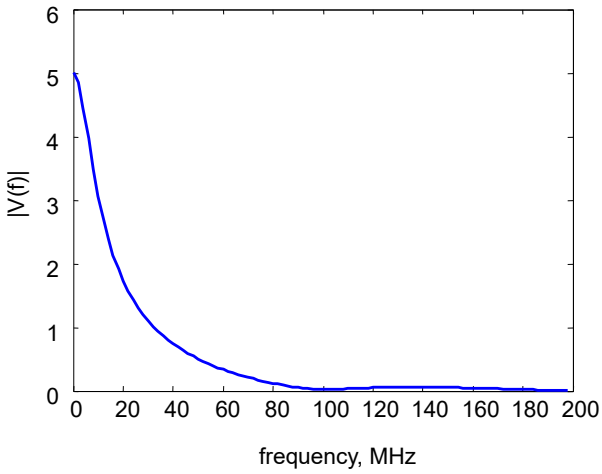
Equivalent voltage spectrum in the frequency domain:

$$V_i(\omega) = \frac{V_\infty\left\{1-\exp\left[-\left(\alpha_1 - i\omega\right)t_0\right]\right\}}{\alpha_1 - i\omega} + \frac{V_\infty\left\{1-\exp\left[i\omega t_0\right]\right\}}{i\omega} - \frac{V_0\exp\left[i\omega t_0\right]}{\alpha_2 - i\omega}$$

A plot of the magnitude of this spectrum at the parameter values listed previous looks like the previous plot:

However, we do not have to use the FFT to compute the frequency spectrum in this case since we can analytically perform the Fourier transform, which is shown here

Pulser Modeling



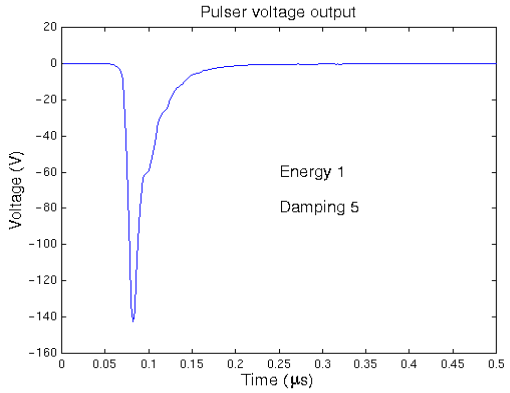
Here is a plot of the analytical Fourier transform which is indistinguishable from the FFT result.

```
>> vff = pulserV(200, .01, 0.2,50,f);
>> plot(f(1:100), abs(vff(1:100)))
>> xlabel('frequency, Hz')
>> ylabel('|V(f)|')


function Vout = pulserV(V0, t0, a1, a2, f)
f = f + eps*(f ==0);
Vinf = V0/(1 -exp(-a1*t0));
arg1 = a1 - i*2*pi*f;
arg2 = a2 - i*2*pi*f;
Vout = Vinf*(1-exp(-arg1*t0))./arg1 + Vinf*(1 -exp(i*2*pi*f*t0))./(i*2*pi*f) ...
   -V0*exp(i*2*pi*f*t0)./arg2;
```

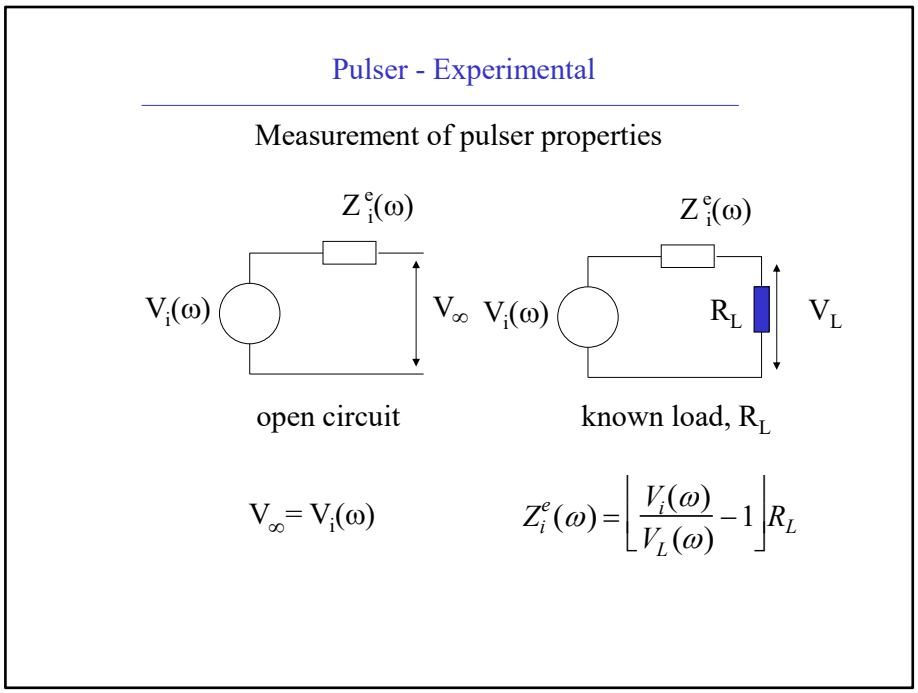
Here is the MATLAB code for plotting this analytical result.

Pulser – Panametrics 5052 PR

Output pulse of Panametrics 5052PR, energy setting = 1, damping setting = 5, open output terminal



Here is the actual open circuit voltage measured for the Panametrics 5052PR pulser/receiver at the indicated energy and damping settings. It looks very much like our simulated four parameter mode. This is also the time domain voltage source in the Thevenin equivalent model.

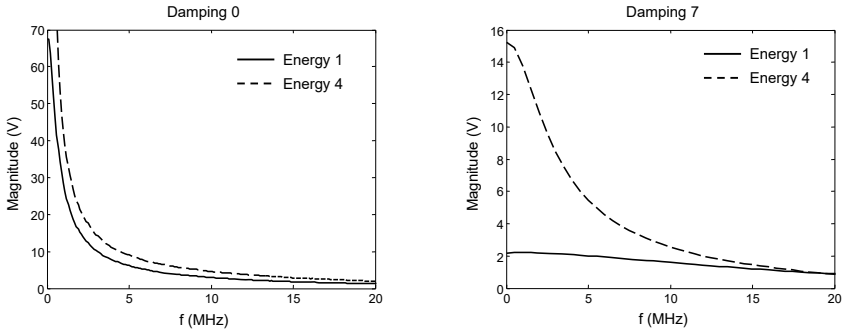


We get the Thevenin equivalent voltage source in a Thevenin equivalent model from an open circuit voltage measurement and we can likewise determine the Thevenin equivalent electrical impedance by measuring the voltage across a known load, such as a resistance, placed across the output.

Pulser – Panametrics 5052PR

Spike Pulser

$$V_i(\omega)$$

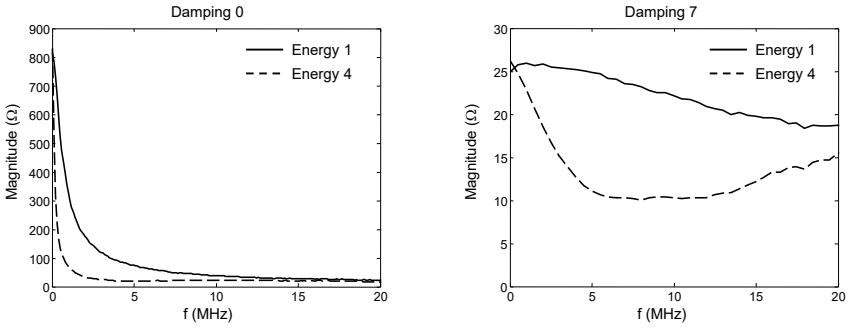


Here are some pulser voltage source measurements in the frequency domain taken at different energy and damping settings

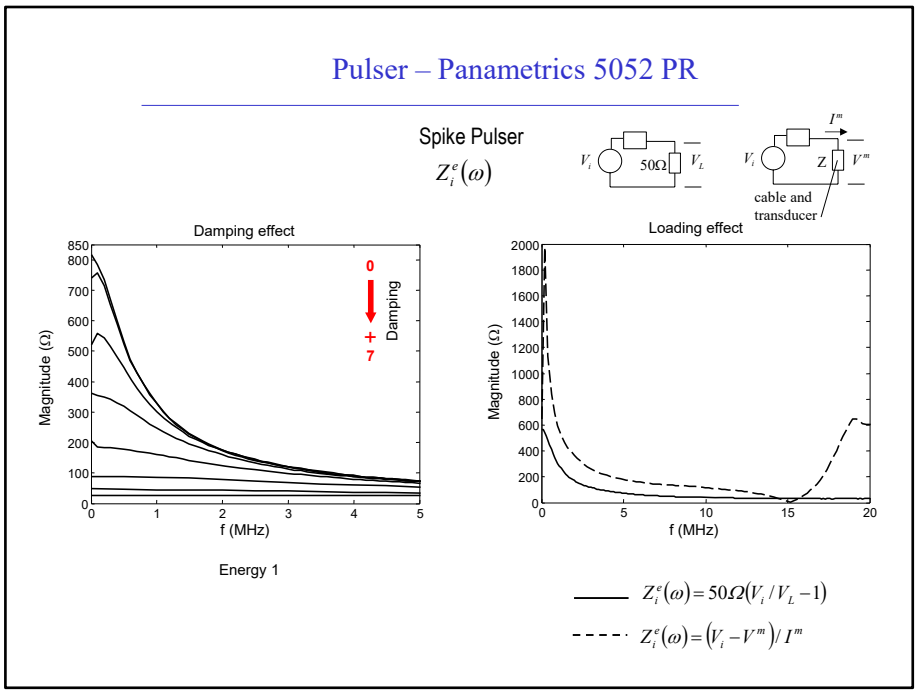
Pulser – Panametrics 5052 PR

Spike Pulser

$$Z_i^e(\omega)$$



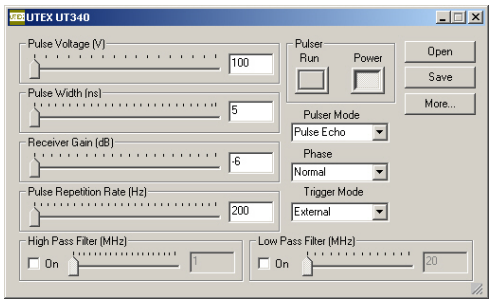
Similarly, here are some impedance measurements taken at different energy and damping settings



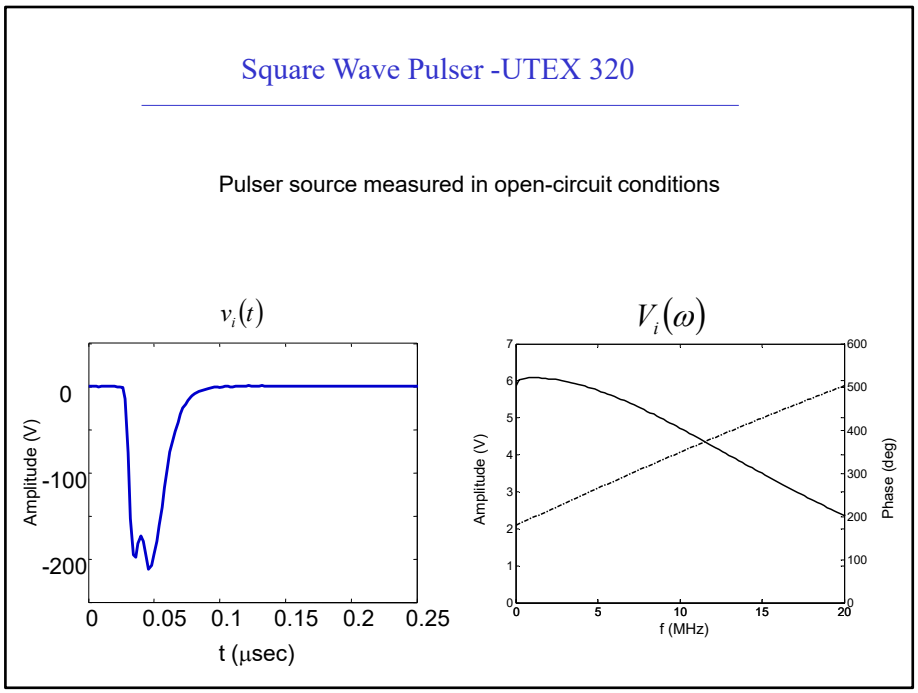
To get a better picture of the effects of the damping setting on the measured Thevenin equivalent impedance, on the left we see the frequency curves over a wide range of the damping values (at an energy 1 setting).

On the right we show the impedance calculated by measuring the voltage, V_L, across a 50 ohm resistor (solid line) at the output of the pulser as well as the impedance calculated when a cable and transducer are attached to the pulser output and both the current and voltage (V^m, I^m) at the output are measured (dashed line). In principle, we should get the same result in either case but we do see some differences depending on the loading of the pulser at its output port. In both cases V_i is the Thevenin equivalent voltage source. Thus, it might be advisable to conduct these measurements under loading conditions actually present in an experiment.

Square Wave Pulser -UTEX 340



Another type of pulser/receiver commonly used in practice contains a square wave pulser such as the UTEX 340. This instrument can be controlled from its front panel or from an equivalent computer interface, as shown. In this case there are three pulser controls consisting of the **pulse repetition rate**, the **pulse width**, and the **pulse voltage**. The gain and high and low pass filter controls are for the receiver section.

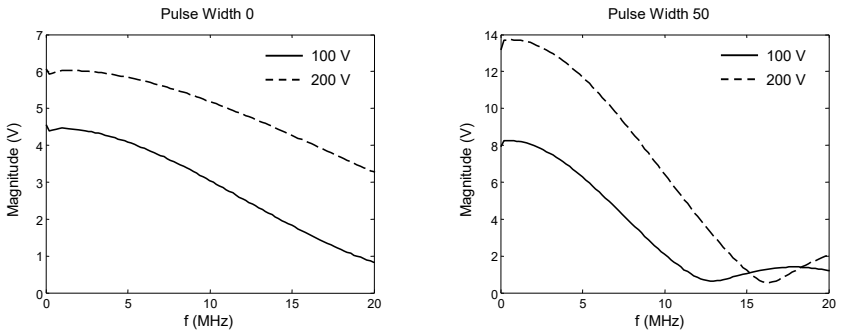


Here is the measured open circuit voltage and its frequency components, showing the square-like nature of the pulse in the time-domain

Square Wave Pulser -UTEX 320

Square wave Pulser

$$V_i(\omega)$$

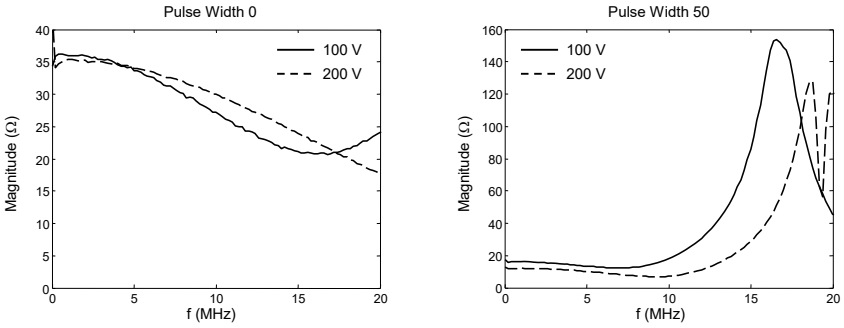


This shows as the pulse width widens the frequency domain response does become less broad.

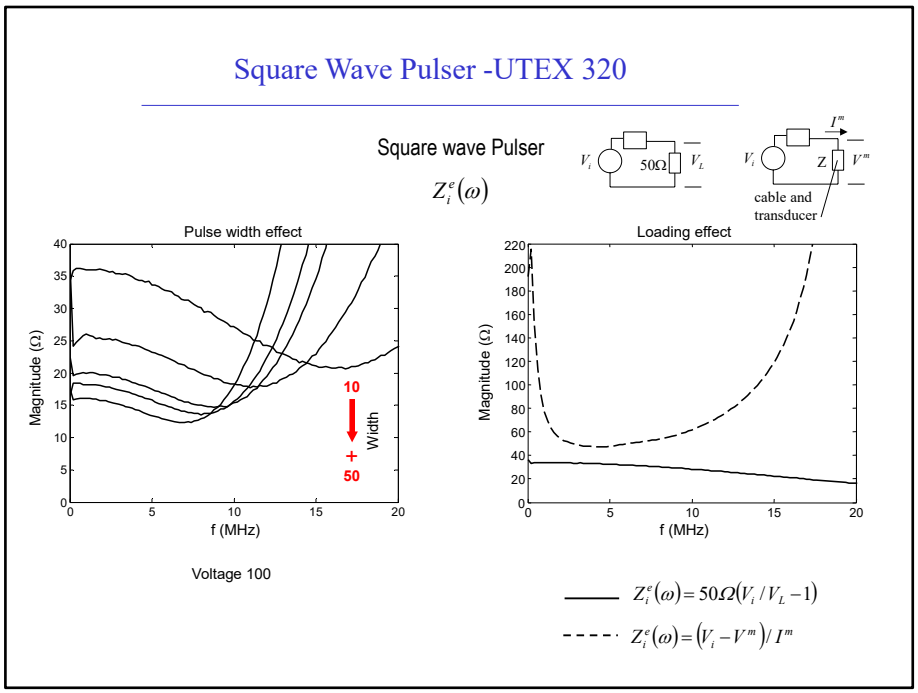
Square Wave Pulser -UTEX 320

Square wave Pulser

$$Z_i^e(\omega)$$

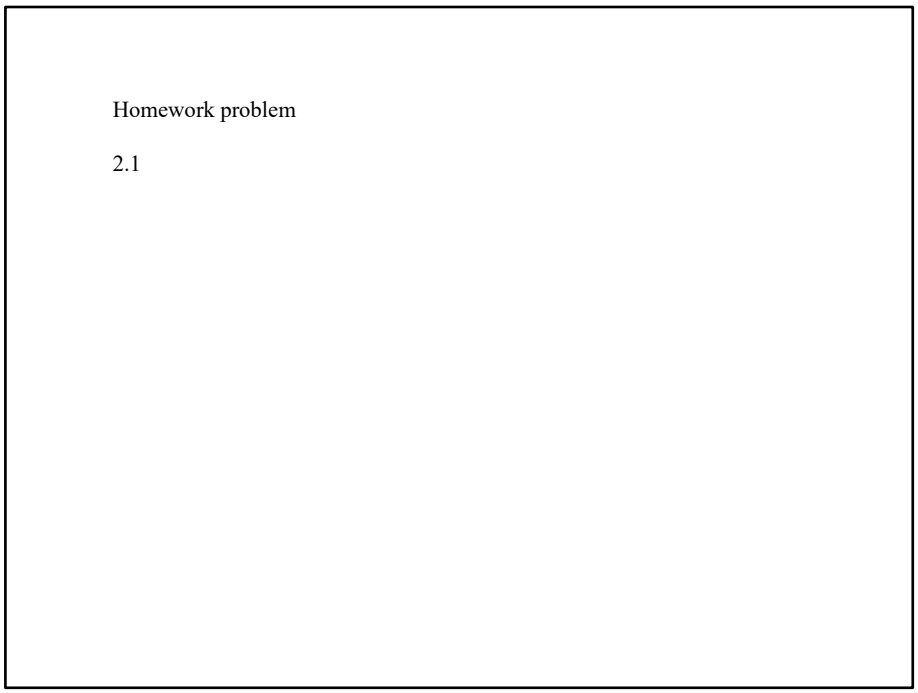


Here are some equivalent impedance measurements under different pulse width and pulse amplitude settings.



On the left we see the effects of changing pulse width settings on the equivalent impedance.

On the right we see the dependency of this impedance on the loading used at the output terminal. Again, the first case (solid line) is where a 50 ohm resistor is placed at the output and the voltage V_L is measured, while the second case (dashed line) is where a cable and transducer are attached an the voltage and current are both measured.



Homework problem from chapter 2.

Linear Electrical and Acoustic Systems
– Some Basic Concepts

Linear systems play an important role in modeling ultrasonic NDE systems so we will discuss some of the basic concepts involved.

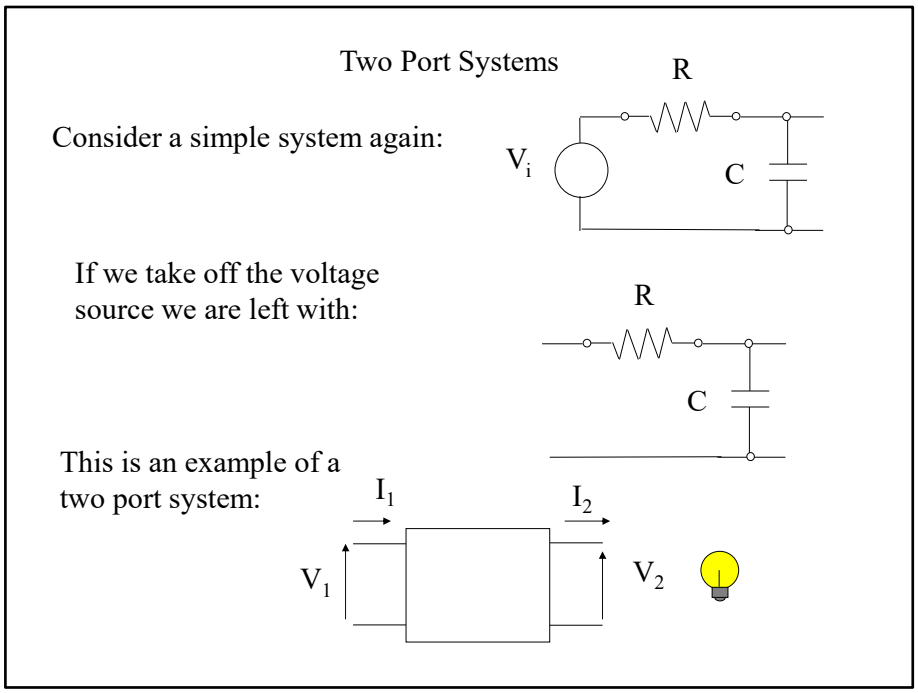
Learning Objectives

Two port systems
- transfer matrices
- impedance matrices
- reciprocity

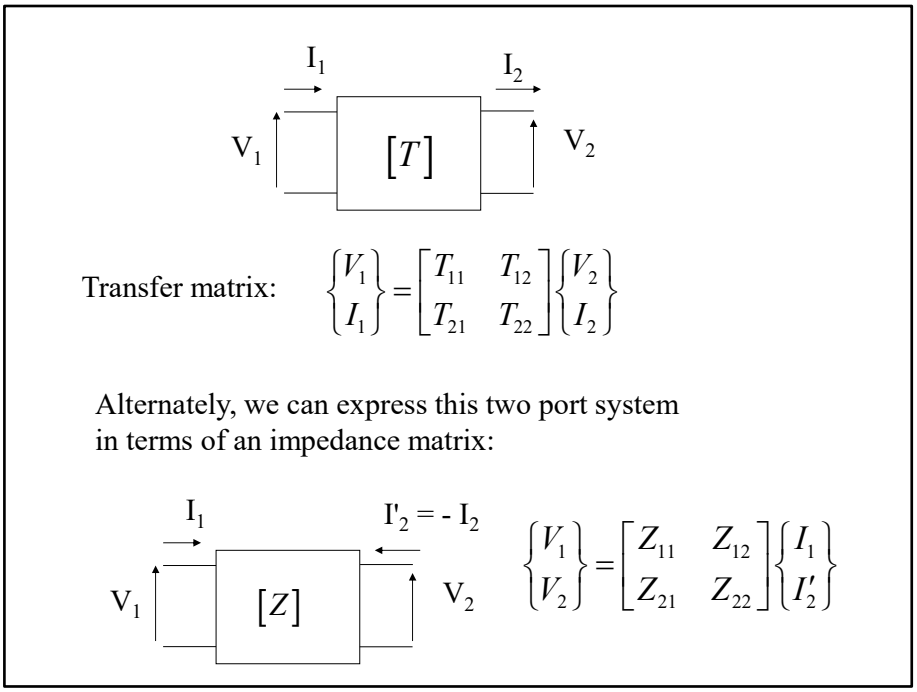
1-D compressional waves in a solid
- equation of motion/constitutive equation
- acoustic transfer matrix of a layer

Single Input-Single Output systems
- LTI systems
- impulse response, transfer functions
- convolution/ deconvolution
- Wiener filter

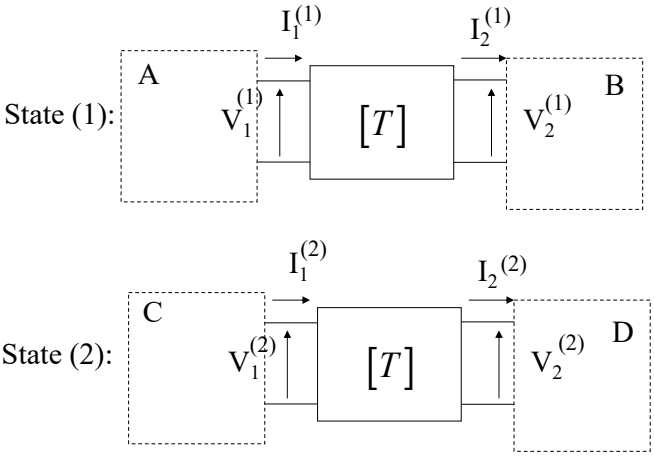
We will see that both two port systems and single input-output systems are important types of systems for modeling ultrasonic systems for both the electrical and acoustic components.



Here we see that an RC circuit with the voltage source removed can be modeled as a **two port electrical system** where we can have a voltage and current acting at both the left and right ends (ports) of the system.



For a linear two port system the voltage and current flowing into the input port can be related to the voltage and current flowing out of the output port through a **2x2 transfer matrix**. Similarly, we can model the two port system in terms of a **2X2 impedance matrix** where the voltages are related to the currents. However, in this case by convention we take the currents as both flowing into the system, as shown.



Reciprocity

$$V_1^{(1)} I_1^{(2)} - V_1^{(2)} I_1^{(1)} = V_2^{(1)} I_2^{(2)} - V_2^{(2)} I_2^{(1)}$$

Consider a two port system that is connected to two sets of other systems, producing what are called states (1) and (2), as shown above. If the two port system is a **reciprocal** system then these voltages and currents in these two states are related through a reciprocity equation.

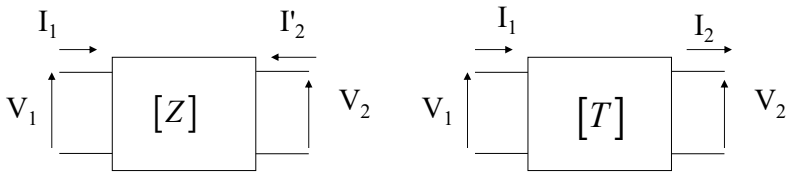
For reciprocal systems, the impedance matrix is symmetric, i.e.

$$Z_{21} = Z_{12}$$

and the determinant of the transfer matrix is equal to one

$$\det[T] = T_{11}T_{22} - T_{12}T_{21} = 1$$

$$\begin{Bmatrix} V_1 \\ V_2 \end{Bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{Bmatrix} I_1 \\ I_2' \end{Bmatrix} \qquad \begin{Bmatrix} V_1 \\ I_1 \end{Bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{Bmatrix} V_2 \\ I_2 \end{Bmatrix}$$



It can be shown that reciprocity implies that the impedance matrix of the two port system is symmetric and that the determinant of the 2x2 transfer matrix is unity.

Relationship between transfer matrix components and impedance matrix components (reciprocal system)

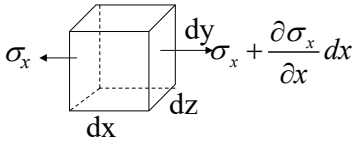
$$T_{11} = \frac{Z_{11}}{Z_{12}}, \quad T_{12} = \frac{\left(Z_{11}Z_{22} - Z_{12}^2\right)}{Z_{12}}$$

$$T_{21} = \frac{1}{Z_{12}}, \quad T_{22} = \frac{Z_{22}}{Z_{12}}$$

Note: $T_{12} \neq T_{21}$

Since the impedance matrix and the transfer matrix both describe the same two port system, they must be related. For a reciprocal system, here is how the transfer matrix components are related to the impedance matrix components.

1-D plane compressional wave in an elastic solid



σ_x …stress

u_x … displacement

$$\left(\sigma_x + \frac{\partial \sigma_x}{\partial x} dx\right) dydz - \sigma_x dydz = \rho dxdydz \frac{\partial^2 u_x}{\partial t^2}$$

$$\frac{\partial \sigma_x}{\partial x} = \rho \frac{\partial^2 u_x}{\partial t^2}$$

Constitutive equation

$$\sigma_x = \frac{E(1-\nu)}{(1+\nu)(1-2\nu)} \frac{\partial u_x}{\partial x}$$

$$= \rho c_P^2 \frac{\partial u_x}{\partial x}$$

compressional (P) wave speed

$$c_P = \sqrt{\frac{E(1-\nu)}{(1+\nu)(1-2\nu)\rho}}$$

Previously, we examined 1-D plane waves traveling in a fluid. Now, consider a 1-D plane compressional wave traveling in a solid where we assume the wave is traveling in the x-direction and the only stress acting in that direction is a normal stress in the x-direction We can again apply Newton's law to relate the stress to the x-displacement. For the solid we have a stress-strain constitutive equation as shown, which can be written in terms of Youngs modulus, E, and Poisson's ratio, v, or equivalently, in terms of the density ,ρ, and the wave speed, c_p , for compressional waves

$$\frac{\partial^2 u_x}{\partial x^2} - \frac{1}{c_P^2}\frac{\partial^2 u_x}{\partial t^2} = 0$$

displacement

$$u_x(x) = A\exp\left[ik_P x - i\omega t\right] + B\exp\left[-ik_P x - i\omega t\right]$$

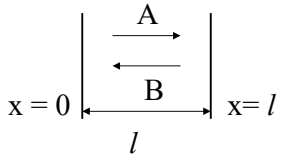
velocity $v_x = -i\omega u_x$

$$v_x(x) = -i\omega A\exp\left[ik_P x - i\omega t\right] - i\omega B\exp\left[-ik_P x - i\omega t\right]$$

stress

$$\sigma_x(x) = i\omega\rho c_P A\exp\left[ik_P x - i\omega t\right] - i\omega\rho c_P B\exp\left[-ik_P x - i\omega t\right]$$

waves in a solid layer



compressive force

$$F_x = -\sigma_x S$$

(minus sign because a positive stress is tensile)

acoustic impedance of the layer

$$Z_0^a = \rho c_P S$$

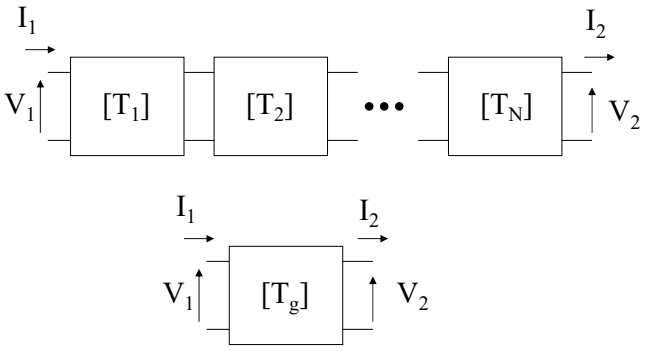
$$\begin{Bmatrix} F_x(0) \\ v_x(0) \end{Bmatrix} = \begin{bmatrix} \cos(k_P l) & -iZ_0^a \sin(k_P l) \\ -i\sin(k_P l)/Z_0^a & \cos(k_P l) \end{bmatrix} \begin{Bmatrix} F_x(l) \\ v_x(l) \end{Bmatrix}$$

The wave speed c_p is indeed the wave speed of this wave in the solid since when we place the constitutive equation into Newton's law we get the wave equation for the displacement where c_p appears as the wave speed present.

Now, consider a combination of harmonic waves traveling in both the plus and minus x-directions. Shown are the expressions for the displacement, velocity and stress due to these waves. Such a combination of waves might exist, for example, in a solid layer of width *l* where we have waves traveling in both directions due to reflections from the ends of the layer. If we write the wave amplitudes A and B in terms of the compressive forces and velocities acting at the ends of the layer then we can relate these compressive forces and velocities at each end (port) to each other, producing the transfer matrix shown for this acoustic two port system, whose elements are functions of wave number, the length, and the acoustic impedance of the layer. The forces involved here are those forces generated over an areas S on the faces of the layer.

Thus, transfer matrices and two port systems can represent both electrical elements and acoustic wave elements of an NDE system.

Equivalent transfer matrices

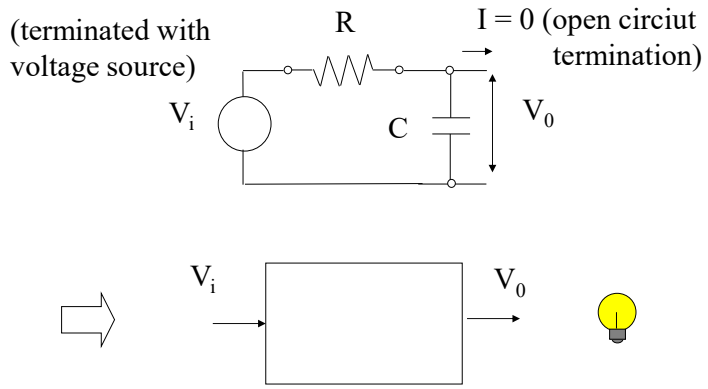


$$\left[T_g\right] = \left[T_1\right]\left[T_2\right] \bullet\bullet\bullet \left[T_N\right]$$

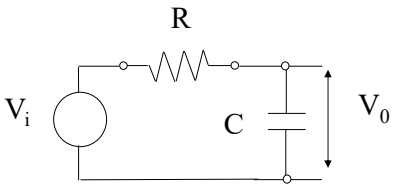
$$\det\left[T_g\right] = \det\left[T_1\right]\det\left[T_2\right] \bullet\bullet\bullet \det\left[T_N\right] = 1$$

If we connect a number of two port systems, then we can replace the entire collection of the system by a single 2x2 transfer matrix that is simply the matrix product of the individual transfer matrices. If the individual systems are reciprocal then the transfer matrix of the collection will also be reciprocal.

When we specify termination conditions at both ports of a two port system, we end up with a system where single inputs and outputs are related:



If we connect the ends of a two port system to either known sources or terminating impedances then the two port system reduces to a **single input, single output system** as shown for example for this RC circuit where a voltage source is placed at one end and the other end is left as an open circuit. In that case we see thesingle input, single output system relates the open circuit output voltage to the input voltage source.



$$V_i(t) - V_0(t) = i(t)R$$

$$i(t) = C\frac{dV_0(t)}{dt}$$

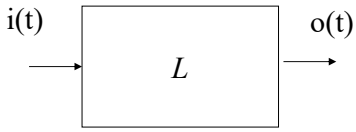
so $$\frac{dV_0(t)}{dt} + \frac{V_0(t)}{RC} = \frac{V_i(t)}{RC}$$

Note: $V_0(t)$ is defined here in terms of $V_i(t)$ only implicitly as the solution of this differential equation, i.e.

$$V_0(t) = L\left[V_i(t)\right] \qquad L \ldots \text{linear operator}$$

If we examine this RC circuit in the time domain we see we can write down a differential equation that gives the open circuit voltage in terms of the voltage source. This is a linear differential equation whose solution can be written symbolically as a linear operator acting on the voltage source to generate the open circuit voltage output.

An important class of these single input-output systems is a linear time-shift invariant (LTI) system



if $o_1(t) = L[i_1(t)]$

$o_2(t) = L[i_2(t)]$

then

$o(t) = L[a_1 i_1(t) + a_2 i_2(t)]$

$= a_1 L[i_1(t)] + a_2 L[i_2(t)]$

linearity

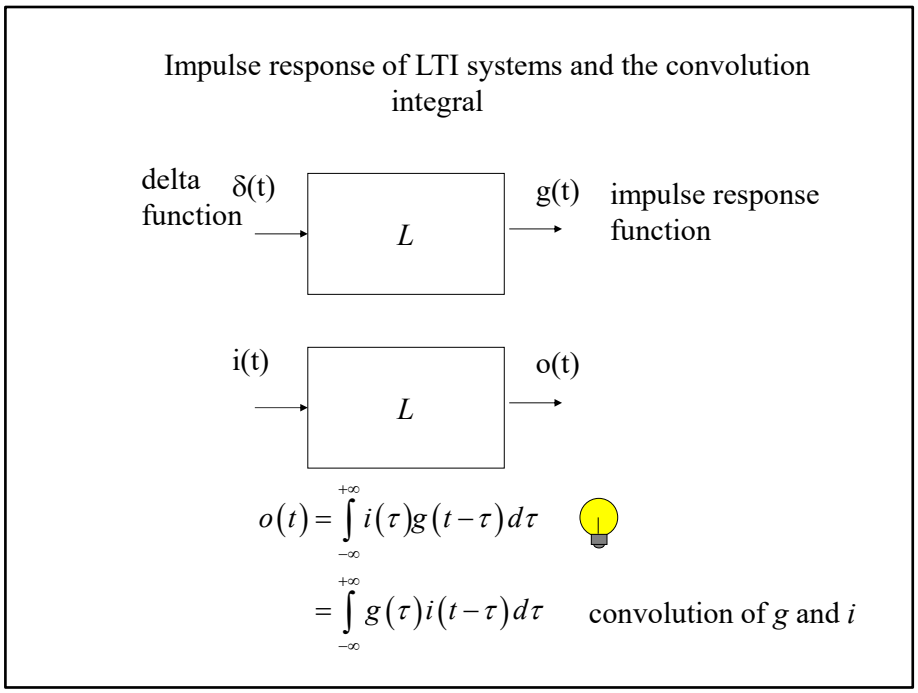
if $o(t) = L[i(t)]$

then

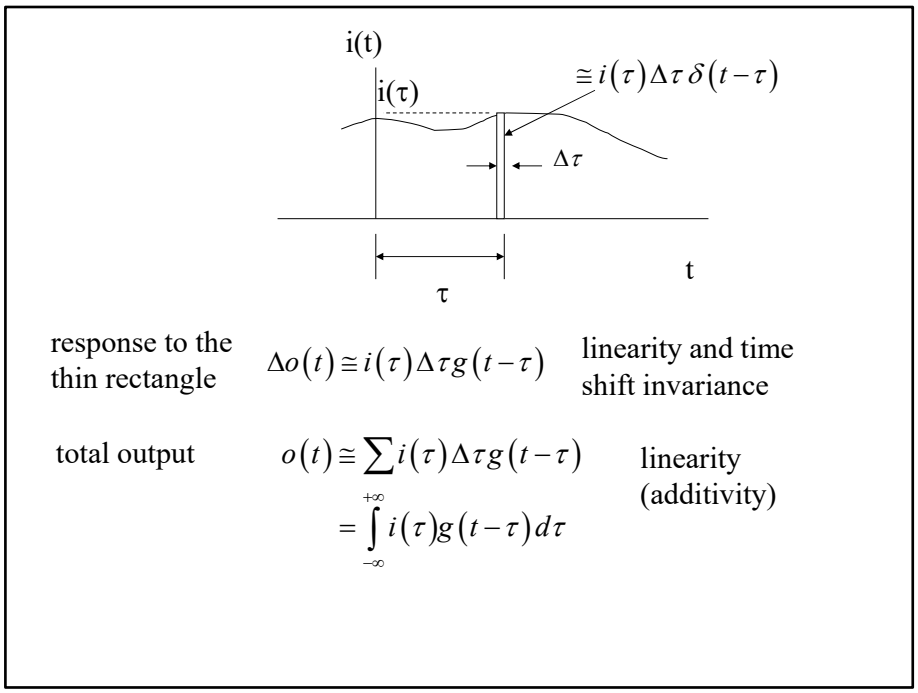
$o(t - t_0) = L[i(t - t_0)]$

time-shift invariance

Since the linear operator transforms the input to the output it can be represented as a single input, single output system as shown. Since the operator here is a linear operator, we say this is a linear system where the inputs and outputs satisfy the linearity conditions shown above. If a time shift at the input produces an identical time shift at the output, then we say this is a **linear time-shift invariant (LTI) system**. Obviously, our RC circuit example is an LTI system. We will assume the components of an ultrasonic NDE system can be represented by LTI systems.



An important characteristic of LTI systems is that if we consider a time domain delta function as the input and let g(t) be the output of the system, called the **impulse response function**, then the response of the LTI system to any input, i(t), can be written as the **convolution integral** of that input with the impulse response function. Thus, the impulse response function completely characterizes the LTI system.



This convolution integral result follows directly by considering an arbitrary input as the sum of different strength delta functions (actually small width rectangles that approximate delta functions) and using the linearity and time shift invariance properties of the LTI system to write the output as a finite sum, which in the limit simply becomes the convolution integral.

If

$$I(\omega) = \int_{-\infty}^{+\infty} i(t) \exp(i\omega t)\, dt$$

$$O(\omega) = \int_{-\infty}^{+\infty} o(t) \exp(i\omega t)\, dt$$

$$G(\omega) = \int_{-\infty}^{+\infty} g(t) \exp(i\omega t)\, dt$$

and $$o(t) = \int_{-\infty}^{+\infty} i(\tau) g(t-\tau)\, d\tau$$

then $$O(\omega) = G(\omega) I(\omega)$$

If we calculate the Fourier transforms of the input function and the impulse response function then one can show that the convolution integral relationship between the time domain inputs and outputs of an LTI system reduces to a multiplication in the frequency domain.

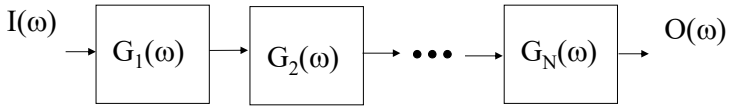
$$o(t) = \int_{-\infty}^{+\infty} i(\tau) g(t-\tau)\, d\tau$$

$$= \int_{-\infty}^{+\infty} g(\tau) i(t-\tau)\, d\tau$$

$$O(\omega) = G(\omega) I(\omega)$$

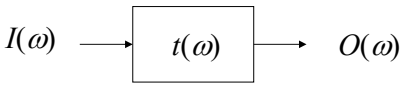
convolution in the frequency domain is just (complex-valued) multiplication

We can write the convolution integral of the input and the impulse response function in the two different equivalent forms shown . However, the input-output relationship is much more conveniently calculated in the frequency domain where we need only to perform a complex-valued multiplication. After the multiplication we can also recover the time domain output signal through an inverse FFT.



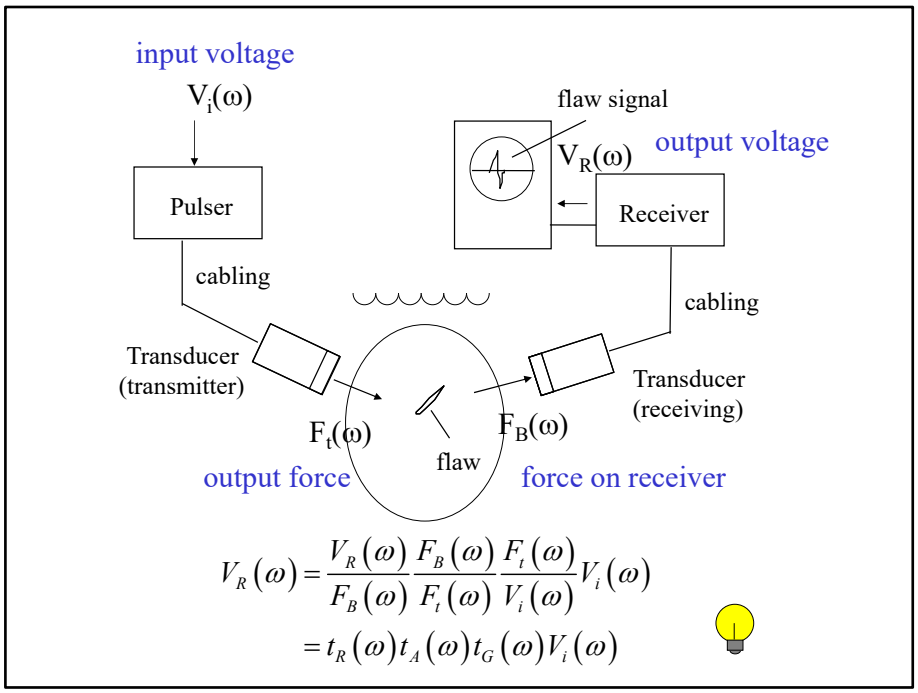
$$O(\omega) = G_1(\omega) G_2(\omega) \cdots G_N(\omega) I(\omega)$$

The frequency components of the impulse response function of an LTI system are also called the transfer function, $t(\omega)$, for the system since this function "transfers" the inputs to the outputs:



$$t(\omega) = \frac{O(\omega)}{I(\omega)}$$

Working in the frequency domain allows us to deal with a cascade of LTI systems that we can replace by a single **transfer function**, which is just the Fourier transform of the impulse response function for the entire system. Formally, the transfer function is just the ratio of the output to the input in the frequency domain.



As an example of such a cascade of LTI systems, consider all the components of an ultrasonic NDE immersion system. We can consider the output to be the frequency components of the received flaw signal and the input to be the Thevenin equivalent voltage source of the pulser. Then we can write the output in terms of the input multiplied by three transfer functions that represent the sound generation process (pulser, cable, sending transducer), the wave propagation and scattering processes between the transducers, and the sound reception process (receiving transducer, cabling, and receiver). We will see later we can model all these transfer functions so we have a complete model of the entire ultrasonic NDE measurement system.

Deconvolution

deconvolution in the frequency domain is just (complex-valued) division … but it must be done with care

$$G(\omega) = \frac{O(\omega)}{I(\omega)}$$

Generally, a Wiener filter is used in ultrasonics applications to desensitize the division process to noise

$$G(\omega) = \frac{O(\omega) I^*(\omega)}{|I(\omega)|^2 + \varepsilon^2 \max\left\{|I(\omega)|^2\right\}}$$

small "noise" constant

( )* = complex conjugate

$$|I(\omega)|^2 = I(\omega) I^*(\omega)$$

In principle we can obtain the transfer function by dividing the frequency components of a measured output by the frequency components of a measured input. However, while convolution, which involves multiplication in the frequency domain, is well behaved, division in the frequency domain (also called **deconvolution**) can be contaminated by noise, rendering the results invalid. To avoid this we use a filter to handle this problem. In ultrasonic NDE we generally use a **Wiener filter** which has a small noise constant, ε, that we must choose.

It is easier to see the Wiener filter if we rewrite the deconvolution in the form

$$G(\omega) = \frac{O(\omega)}{I(\omega)} \frac{|I(\omega)|^2}{|I(\omega)|^2 + \varepsilon^2 \max\left\{|I(\omega)|^2\right\}}$$

$$= \frac{O(\omega)}{I(\omega)} W(\omega) \qquad (1)$$

where $$W(\omega) = \frac{|I(\omega)|^2}{|I(\omega)|^2 + \varepsilon^2 \max\left\{|I(\omega)|^2\right\}}$$

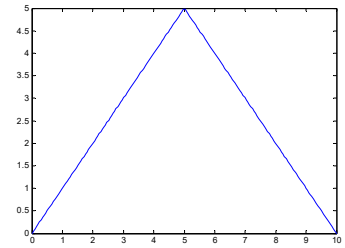
Wiener filter

We can see the effects of the Wiener filter more clearly by writing the deconvolution as a product of the ordinary division by a function W that represents the filter. We will show the behavior of W on the next slide. **However, note that we must not implement the deconvolution in the form of Eq. (1) above but must use the original form shown on the previous page**.

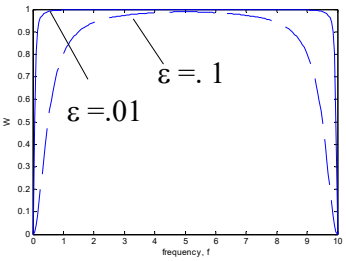
MATLAB example showing effects of choice of ε

```
>> f = linspace(0, 10, 200);
>> I= f.*(f<5) +(10-f).*(f>5);
>> plot(f, I)
>> e =.01;
>> W = I.^2./(I.^2 + e^2*max(I.^2));
>> plot(f, W)
>> hold on
>> e= .1;
>> W = I.^2./(I.^2 + e^2*max(I.^2));
>> plot(f, W,'--')
>> xlabel(' frequency, f')
>> ylabel(' W')
>> hold off
```

I

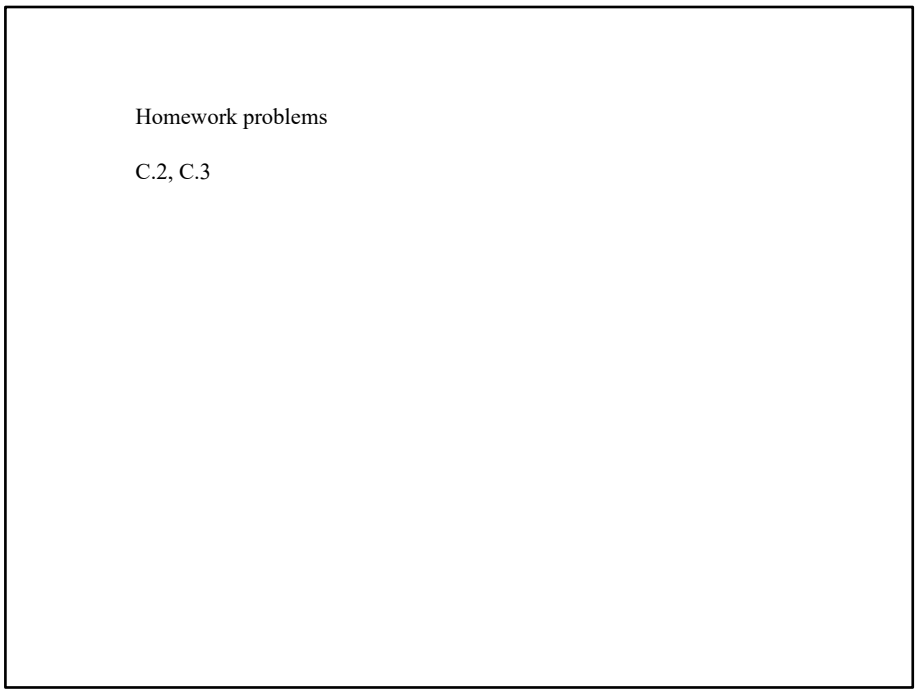


W

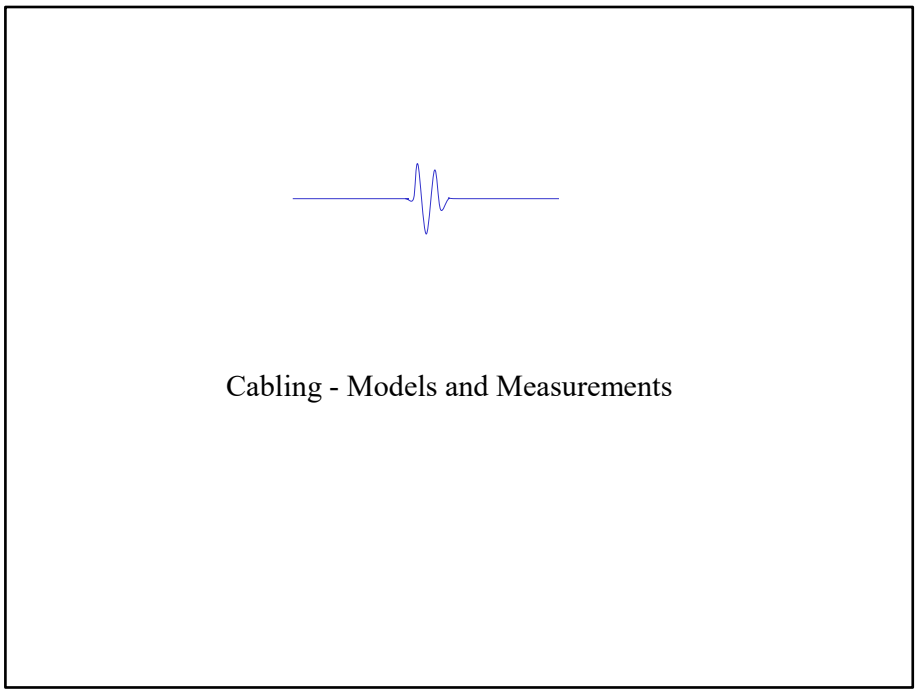


f

Here is an example where we model the input function as a simple triangular function. We see that the Wiener filter W for small values of the constant e is near unity so the Wiener filter does not modify the ordinary division process except at the ends where the input function becomes small. At places where the input is small, noise can contaminate the deconvolution result so the Wiener filter prevents those places from dominating the result.



Homework problems from Appendix C.



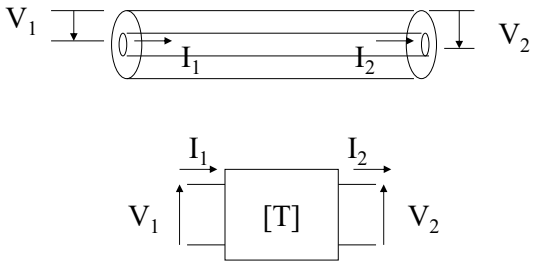
Cabling can play a role in the measurement process for an ultrasonic NDE system so we will discuss the effects of cabling here in some detail.

Learning Objectives

Cable transfer matrix model
Equivalent circuit model
Effects of termination conditions
Measurement of transfer matrix elements
Reciprocity check

We will model a cable as a two port system and show the effects of having different termination conditions. We will describe how we can measure the transfer elements of the cable directly and check to see if it acts as a reciprocal system.

Cable – transfer matrix model

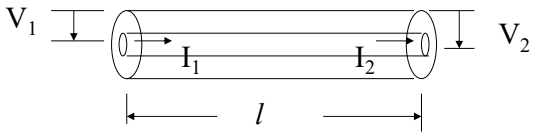


Transfer matrix from a simple 1-D transmission line model

$$\begin{Bmatrix} V_1 \\ I_1 \end{Bmatrix} = \begin{bmatrix} \cos(k_c l) & -iZ_0^e \sin(k_c l) \\ -i\sin(k_c l)/Z_0^e & \cos(k_c l) \end{bmatrix} \begin{Bmatrix} V_2 \\ I_2 \end{Bmatrix}$$

At a fundamental level we can consider modeling the propagating electric and magnetic fields in a coaxial cable. However, we can use a simpler transmission line model where we consider the cable as a two port system that relates the voltages and currents at its ends though a 2x2 transfer matrix of components that depends on an electrical impedance of the cable, its length, and the wavenumber (or frequency).

Cable – transfer matrix model

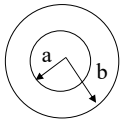


l = length of cable

$k_c = \omega/c$ = wave number (c = wave speed)

Z_0^e = electrical impedance

For an ideal coaxial cable: $Z_0^e = \frac{1}{2\pi}\sqrt{\frac{\mu}{\varepsilon}}\ln\left(\frac{b}{a}\right)$



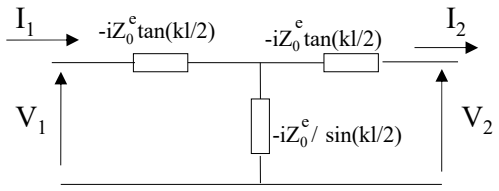
ε … permittivity

μ … permeability

For an ideal circular coaxial cable we can write down an explicit expression for the impedance in terms of the radii of the inner and outer conductors and the electromagnetic properties of the material between those conductors.

Cable – transfer matrix model

Equivalent circuit model of the transmission line model

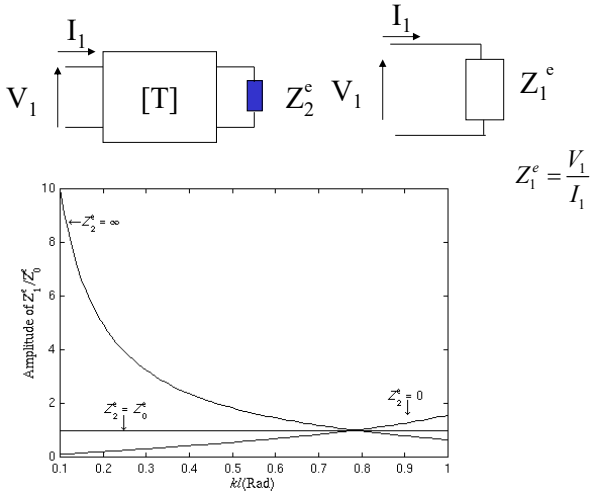


Other equivalent circuits are also used

We can also represent the cable as two port system involving equivalent circuit components. The “T” type of equivalent circuit shown above is a popular choice but there are other equivalent circuits that can be used as well.

Cable – transfer matrix model

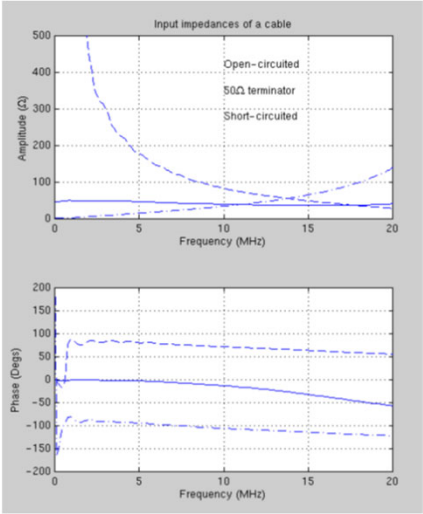
Effects of different termination conditions



If we consider a cable that is terminated at one end, then we can characterize the terminated cable as an equivalent impedance, and by examining the ratio of the input voltage to the input current we can evaluate that equivalent impedance. Show are three cases where the cable is terminated by an impedance equal to the characteristic impedance of the cable and where the termination is either open circuit or closed circuit.

Cable – transfer matrix model

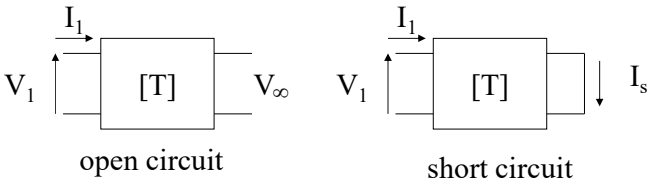
Cable impedance measurements under different terminations



Here are some actual measurements of the magnitude and phase of the equivalent impedance of a 50 ohm coaxial cable when it is terminated by a 50 ohm resistor or is under open or closed circuit conditions. The equivalent impedance is measured by taking the ratio of the measured input voltage to the measured input current.

Cable – transfer matrix model

Measuring the transfer matrix components



open circuit

short circuit

$T_{11} = V_1 / V_\infty$

$T_{21} = I_1 / V_\infty$

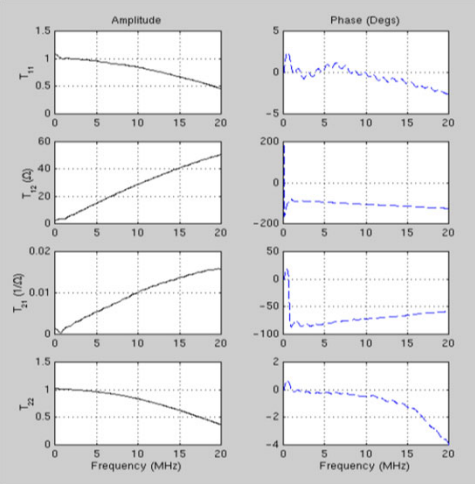
$T_{12} = V_1 / I_s$

$T_{22} = I_1 / I_s$

By making various voltage and current measurements under different termination conditions we can also measure the individual elements of the 2x2 transfer matrix that characterizes the cable as a two port system.

Cable – transfer matrix model

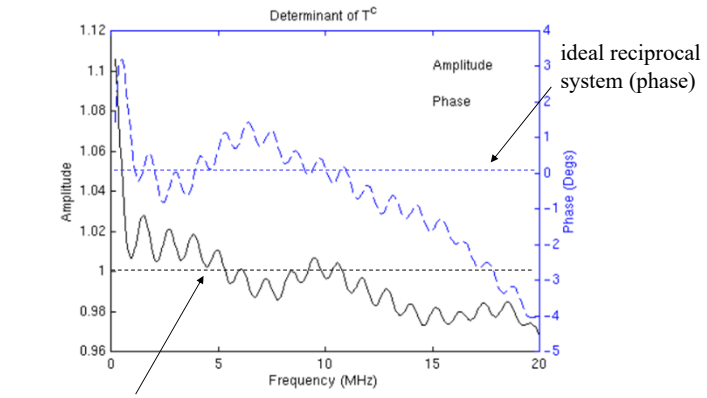
Measured cable transfer components



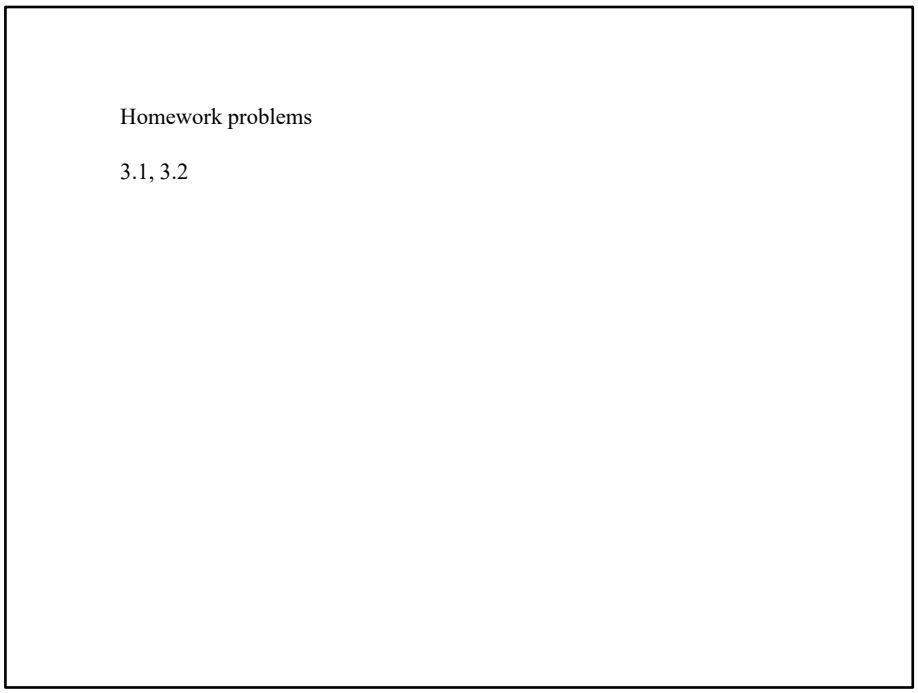
Here are the results of those measurements. It can be seen that the magnitudes of the T11 and T22 components have a cosine like behavior while the magnitudes of the T12 and T21 have a sine like behavior, consistent with the transmission line model. The phases also are consistent with that model

Cable – transfer matrix model

Reciprocity Check



If we evaluate the determinant of the 2x2 transfer matrix we do get nearly a real value of unity for all frequencies, demonstrating the cable indeed is reciprocal.



Homework problems from Chapter 3.

Cable – transfer matrix model

Staelin, D.H., Morgenthaler, A.W., and J. Kong, Electromagnetic Waves,Prentice-Hall, 1994.

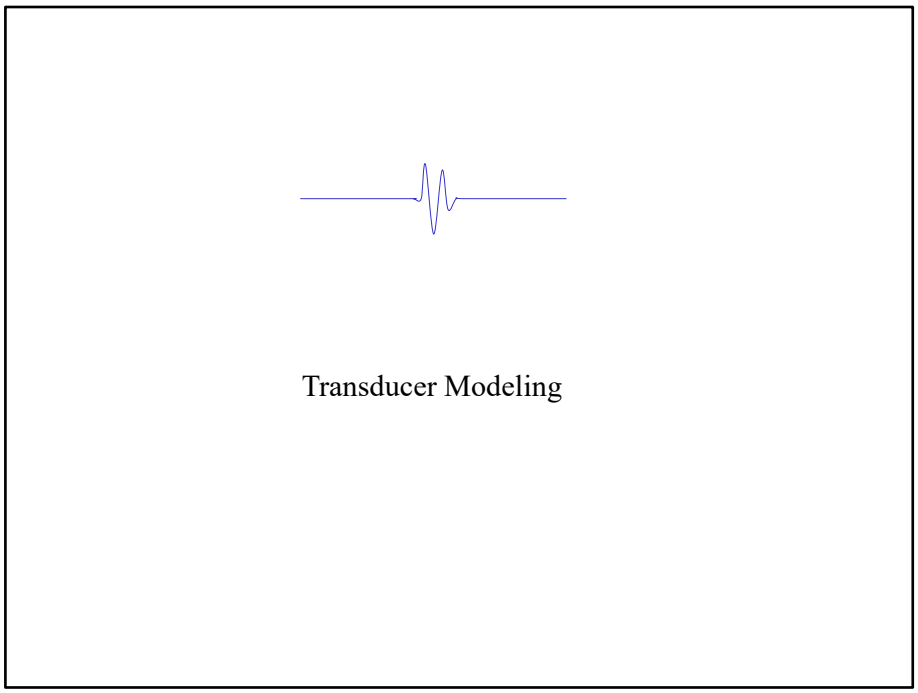
Balanis, C.A., Advanced Engineering Electromagnetics, John Wiley, 1989.

Bladel, J.V., Electromagnetic Fields, Hemisphere Publuishing Co., 1985.

Pozar, D.M., Microwave Engineering, John Wiley, 1998.

Karmel, P.R., Colef, G.D., and R.L. Camisa, Introduction to Electromagnetic and Microwave Engineering, John Wiley, 1998.

Cables are discussed in a variety of EE texts. Here are some representative ones.



Transducers are obviously a very important part of any ultrasonic system. Here we examine models of a transducer when it is used to generate ultrasound.

Learning Objectives

Two port and three port transducer models
- Sittig model
- Mason model
- KLM Model

Acoustic radiation impedance

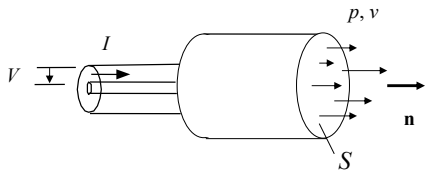
Transducer sensitivity, impedance

The sound generation process

There are a variety of transducer models available that we will examine. We will see that three key parameters associated with the transducer are the **acoustic radiation impedance**, its **electrical impedance** and its **sensitivity**. In this section we will also combine our pulser, cabling and transducer models into a complete representation of the **sound generation process** in an ultrasonic setup.

Transducer model - fields

Consider a P-wave immersion transducer:



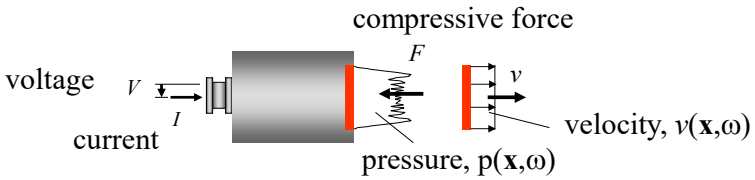
pressure, velocity fields

Reciprocity :

$$V^{(2)}I^{(1)} - V^{(1)}I^{(2)} = \int_S \left(p^{(2)}\mathbf{v}^{(1)} - p^{(1)}\mathbf{v}^{(2)} \right) \cdot \mathbf{n}\, dS \qquad (1)$$

We can consider a sending immersion transducer as a “mixed” model where we take the electrical inputs as lumped parameters of voltage and current but take the outputs as the pressure and velocity fields on the face of the transducer in contact with the fluid. If we assume the transducer is a reciprocal device then we can relate these parameters through the reciprocity relation of Eq. (1)

Transducer model- 'lumped' parameters



For piston behavior

$$\mathbf{v}(\mathbf{x},\omega) = v(\omega)\mathbf{n}$$

$$F(\omega) = \int_S p(\mathbf{x},\omega)\,dS$$

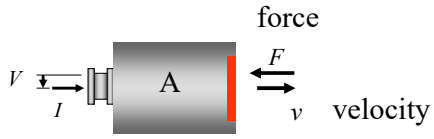
$$\int p(\mathbf{x},\omega)\,\mathbf{v}(\mathbf{x},\omega)\cdot\mathbf{n}\,dS = F(\omega)v(\omega)$$

so reciprocity becomes:

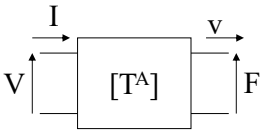
$$V^{(2)}I^{(1)} - V^{(1)}I^{(2)} = F^{(2)}v^{(1)} - F^{(1)}v^{(2)}$$

If we assume the velocity of the transducer face is a constant, then this is a "piston" transducer model and we can replace the integrals of the pressure fields in the reciprocity relations by compressive force terms and write the reciprocity relationship entirely in terms of lumped quantities of (voltage, current) and (force, velocity)

Transducer model - transfer matrix



2-Port Transducer Model

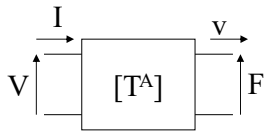


$$\begin{Bmatrix} V \\ I \end{Bmatrix} = \begin{bmatrix} T_{11}^A & T_{12}^A \\ T_{21}^A & T_{22}^A \end{bmatrix} \begin{Bmatrix} F \\ v \end{Bmatrix} \quad , \quad \det\left[T^A\right] = 1$$

From reciprocity

If the transducer is a linear, reciprocal device we can thus relate the lumped inputs and outputs through a two port 2x2 transfer matrix whose determinant must be equal to one.

Transducer model - transfer matrix



Sittig model: $\left[T^A\right] = \left[T_e^A\right]\left[T_a^A\right]$

$$\left[T_e^A\right] = \begin{bmatrix} 1/n & n/i\omega C_o \\ -i\omega C_o & 0 \end{bmatrix}$$

$$\left[T_a^A\right] = \frac{1}{Z_b^a - iZ_0^a \tan(kd/2)} \begin{bmatrix} Z_b^a + iZ_0^a \cot(kd) & \left(Z_0^a\right)^2 + iZ_0^a Z_b^a \cot(kd) \\ 1 & Z_b^a - 2iZ_0^a \tan(kd/2) \end{bmatrix}$$

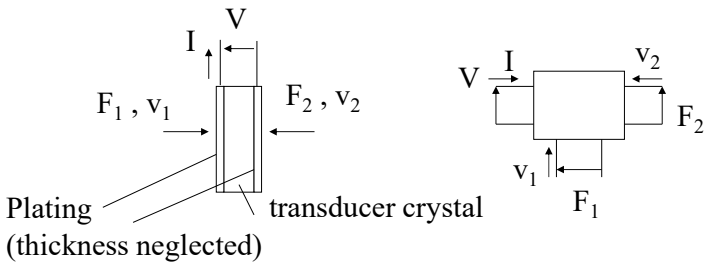
There is an explicit model of the transducer 2x2 transfer matrix called the **Sittig model**, where the overall transfer matrix is written in terms of the product of two separate transfer matrices, as shown.

Transducer model - transfer matrix

$k = \omega / v_0$	wave number of piezoelectric plate
$v_0 = \sqrt{c_{33}^D / \rho_p}$	wave speed of the plate, defined in terms of:
c_{33}^D	plate elastic constant at constant flux density
ρ_p	plate density
$n = h_{33} C_0$	constant, defined in terms of:
h_{33}	plate stiffness
$C_0 = S / \beta_{ss}^S d$	clamped capacitance, defined in terms of:
S, d	plate area, thickness
β_{33}^D	plate dielectric impermeability at constant strain
$Z_0^a = \rho_p v_0 S$	plate acoustic impedance
Z_b^a	backing acoustic impedance

There are many parameters in the Sittig model since the transducer is inherently a complex electromechanical device.

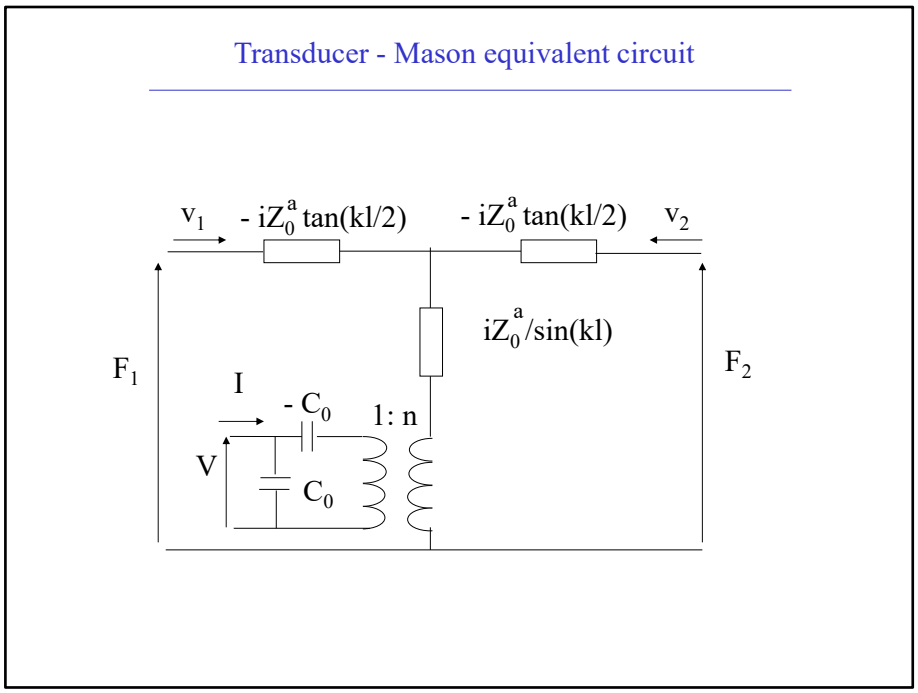
Transducer - three port model



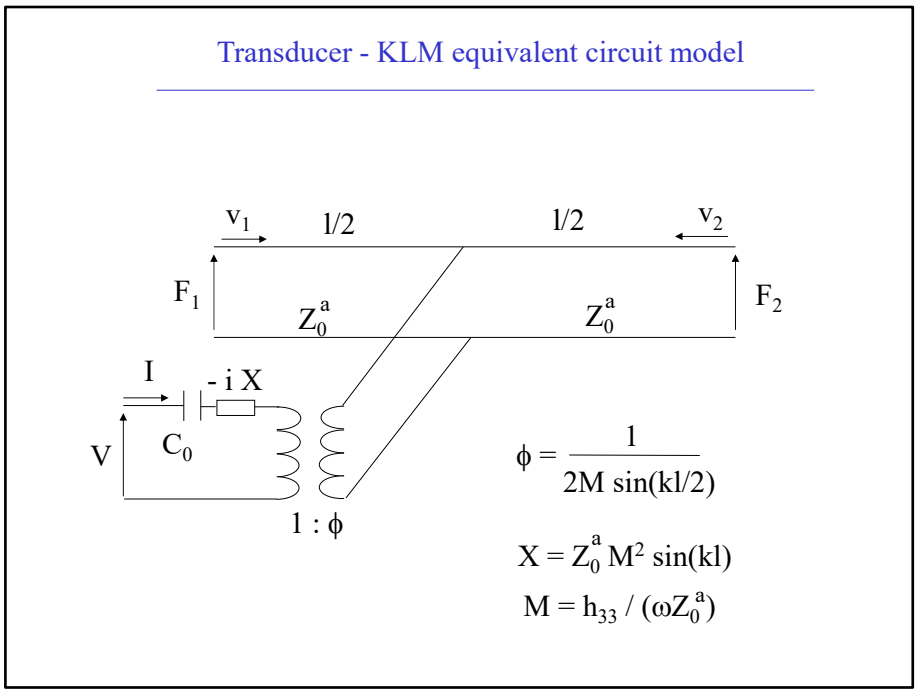
3x3 impedance matrix

$$\begin{Bmatrix} F_1 \\ F_2 \\ V \end{Bmatrix} = i \begin{bmatrix} Z_0^a \cot(kl) & Z_0^a / \sin(kl) & h_{33} / \omega \\ Z_0^a / \sin(kl) & Z_0^a \cot(kl) & h_{33} / \omega \\ h_{33} / \omega & h_{33} / \omega & 1 / \omega C_0 \end{bmatrix} \begin{Bmatrix} v_1 \\ v_2 \\ I \end{Bmatrix}$$

The Sittig model assumes the piezoelectric crystal in the transducer has some material backing on its inside face. If we do not include such backing in a transducer model then we can consider the transducer as a **three port system** where we have the driving electrical port and acoustic ports on both the front and back faces of the crystal. For such a three port system we can model the system through a 3x3 impedance matrix. Such a system is very useful when we want to design a transducer with specific characteristics as we can also change the backing on the crystal as part of the design.



One frequently used equivalent circuit model of such a three port system is the **Mason model** shown here. Some designers do not like this model because it includes a non-physical negative capacitance term.

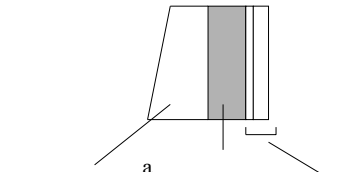


Another equivalent three port circuit model of a transducer is the **KLM model**.

Sittig model with crystal facing layers

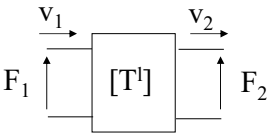
Commercial transducer:

$[T^A] = [T^A_e]\,[T^A_a]\,[T^l]\,\ldots$



Backing (Z_b^a) crystal facing layers (epoxy bonding, wear plate, etc.)

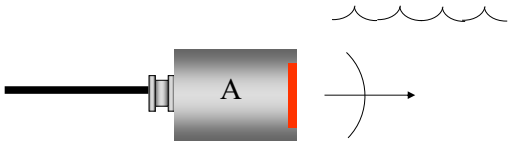
Acoustic layer:



$$\begin{Bmatrix} F_1 \\ v_1 \end{Bmatrix} = \begin{bmatrix} \cos(k_a l_a) & -iZ_0^a \sin(k_a l_a) \\ -i\sin(k_a l_a)/Z_0^a & \cos(k_a l_a) \end{bmatrix} \begin{Bmatrix} F_2 \\ v_2 \end{Bmatrix}$$

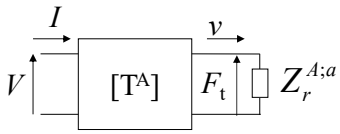
The Sittig, Mason, and KLM model have all been successfully used for designing ultrasonic transducers. Because the Sittig model involves 2x2 transfer matrices, it is particularly easy to add elements such as wear plates since we have seen that such acoustic layers can be characterized as 2x2 acoustic transfer matrices that can simply be concatenated through multiplication with the other transfer matrices.

Transducer – radiation into a fluid



At the acoustic port the force and velocity parameters are not independent. We can write $F_t(\omega) = Z_r^{A;a}(\omega)v(\omega)$

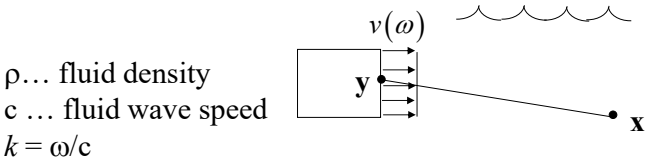
$Z_r^{A;a}$ … acoustic radiation impedance (a "lumped"parameter that depends on the velocity and pressure distribution at the acoustic port, the port geometry, and the fluid properties)



When an immersion transducer is used in practice, its acoustic output port is always inherently terminated, i.e. the output compressive force and the velocity are related to one another through an **acoustic radiation impedance**. This acoustic impedance is a function of the acoustic waves radiated by the transducer so we will have to discuss some elements of those acoustic waves, which we will do here briefly. Later, we will examine those waves in more detail and justify some of the results that are simply given here.

Acoustic radiation impedance

Rayleigh-Sommerfeld integral model of radiation of waves into a fluid by a piston transducer



ρ… fluid density
c … fluid wave speed
$k = \omega/c$

pressure $$p(\mathbf{x},\omega) = \frac{-i\omega\rho v(\omega)}{2\pi}\int_S \frac{\exp(ikr)}{r}\,dS(\mathbf{y})$$

$$r = |\mathbf{x}-\mathbf{y}|$$

$$Z_r^a(\omega) = \frac{\int_s p(\mathbf{x},\omega)\,dS}{v(\omega)} = \frac{-i\omega\rho}{2\pi}\int_S \left\{\int_S \frac{\exp(ikr)}{r}\,dS(\mathbf{y})\right\} dS(\mathbf{x})$$

As we will see later we can model the pressure waves generated in a fluid by a piston transducer in terms of a **Rayleigh-Sommerfeld integral (also called a Rayleigh integral by some authors)** over the face of the transducer. Since the compressive force generated is an integral of this pressure over the face of the transducer, we can write an explicit expression for the acoustic radiation impedance in terms of two surface integrals.

Acoustic radiation impedance

Greenspan, 1979: showed that for a circular piston transducer of radius a the acoustic radiation impedance obtained from the Rayleigh-Sommerfeld model could be found explicitly in the form

$$Z_r^{A;a} / \rho c S_A = 1 - \left[J_1(2ka) - iS_1(2ka) \right] / ka$$

J_1 … Bessel function

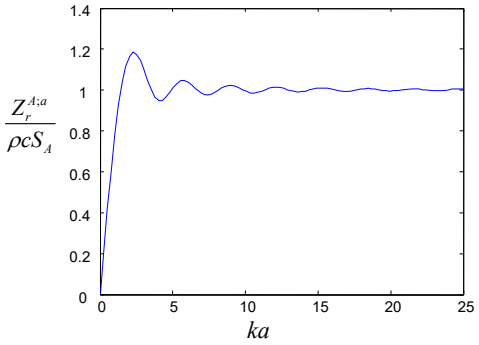
S_1 … Struve function

$S_A = \pi a^2$

However, we do not have to do those integrals directly since Greenspan has shown that for a circular piston transducer we can express those integrals in terms of a **Bessel function** and a **Struve function**, which are two well-known special functions whose numerical evaluation is straightforward.

Acoustic radiation impedance

Greenspan model of a circular piston transducer




```
>> ka=linspace(0, 25, 100);
>> ka = ka + eps*( ka ==0);
>> Z = 1 -(besselj(1,2*ka)-i*struve(2*ka))./ka;
>> plot(ka, abs(Z))
```

Here is a plot of the normalized radiation impedance of a circular piston transducer of radius a, as a function of the nondimensional wavenumber (frequency). We see at larger ka values the acoustic radiation impedance becomes nearly a constant. Here S_A is the area of the face of the transducer. Shown also is the MATLAB code used for evaluation of this plot. The Bessel function can be calculated with a built-in MATLAB function besselj. The struve function will be given on the next slide.

Acoustic radiation impedance

```
function y = struve(z)
num = length(z);
y=zeros(1,num);
for k = 1:num
y(k) = quadl(@struve_arg, 0, 1, [ ],[ ], z(k));
end
```

```
function y = struve_arg(x, z)
y = (4./pi).*z.*x.^2.*sin(z.*(1-x.^2)).*sqrt(2-x.^2);
```

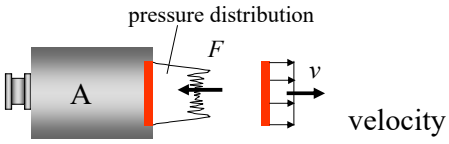
↓ this uses

$$S_1(z) = \frac{2z}{\pi}\int_0^1 \sqrt{1-t^2}\,\sin(zt)\,dt \quad t = 1-x^2$$

$$= \frac{4z}{\pi}\int_0^1 x^2 \sin\left[z\left(1-x^2\right)\right]\sqrt{2-x^2}\,dx$$

Here is the evaluation of the Struve function in MATLAB.

Acoustic radiation impedance



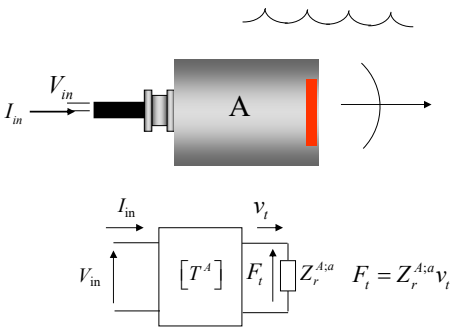
Most NDE transducers operate at high frequencies (ka >> 1). At such high frequencies, if we can assume piston behavior, for any shaped transducer it can be shown that

$$Z_r^{A;a} \cong \rho c S_A$$

density, wave speed, area

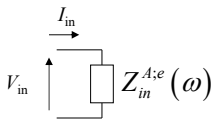
Most NDE transducers operate at high frequencies and we can show that in fact the acoustic radiation impedance for any such high frequency transducer (not just a circular one) is equal to the acoustic plane wave impedance, which is the product of the density of the fluid times its wave speed times the area of the face of the transducer. Since this acoustic radiation impedance is a known constant, when dealing with a sending transducer we can always treat it as a terminated system where we know the value of that acoustic termination.

Sensitivity, Impedance



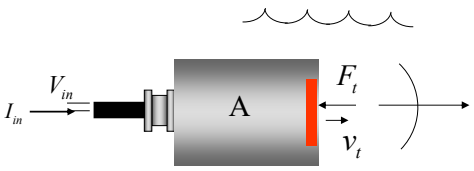
the electrical characteristics of the transducer can be completely described by its input impedance:

$$Z_{in}^{A;e} = \frac{V_{in}}{I_{in}} = \frac{Z_r^{A;a} T_{11}^A + T_{12}^A}{Z_r^{A;a} T_{21}^A + T_{22}^A}$$



Unlike a cable, it is very difficult to measure the 2x2 transfer matrix that characterizes a transducer because one port is an acoustic port where it is not simple to measure either force or velocity. However, because the transducer is always terminated with a known acoustic impedance, as shown above, we will see that we do not have to measure all the transfer components directly. For example, we can easily measure the voltage and current at the input electrical port, whose ratio gives us the **transducer electrical impedance**. This electrical impedance tells us how the transducer electrically terminates the cable it is connected to.

Sensitivity, Impedance



to describe the conversion of electrical signals into acoustic signals, we could use the transducer's sensitivity, S_{OI}, where $S_{OI} = \frac{O}{I}$

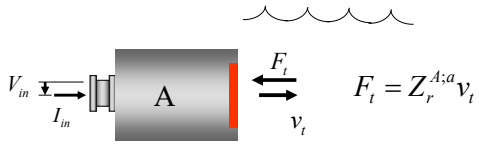
O … an output (force or velocity)
I … an input (voltage or current)

The particular sensitivity we will use is:

$$S_{vI}^{A} \equiv \frac{v_t}{I_{in}} = \frac{1}{Z_r^{A;a} T_{21}^{A} + T_{22}^{A}}$$

If, in addition to knowing the electrical termination (impedance) characteristics of the transducer, we also have a measure of how the electrical signals are converted into output force or velocity, then we will have described the ability of the transducer to generate its acoustic output from the electrical signals. We can, for example, specify a **transducer sensitivity** as a ratio of an output (force or velocity) to an input (voltage or current). The specific sensitivity we will use is the ratio of the output velocity to the input current. Shown above is this sensitivity in terms of the transfer matrix components.

Sensitivity, Impedance



$$F_t = Z_r^{A;a} v_t$$

All the other sensitivities can be found from this sensitivity if the transducer electrical impedance and acoustic radiation impedance are known:

$$S_{vI}^{A} = \frac{v_t}{I_{in}}$$

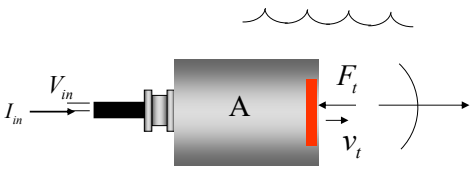
$$S_{FI}^{A} = \frac{F_t}{I_{in}} = Z_r^{A;a} S_{vI}^{A}$$

$$S_{vV}^{A} = \frac{v_t}{V_{in}} = S_{vI}^{A} / Z_{in}^{A;e}$$

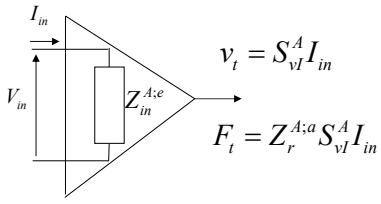
$$S_{FV}^{A} = \frac{F_t}{V_{in}} = Z_r^{A;a} S_{vI}^{A} / Z_{in}^{A;e}$$

Our choice of sensitivity was arbitrary but any other transducer sensitivity we might want to define can be obtained from this one (and the electrical and acoustic impedances) so this particular choice is all we need.

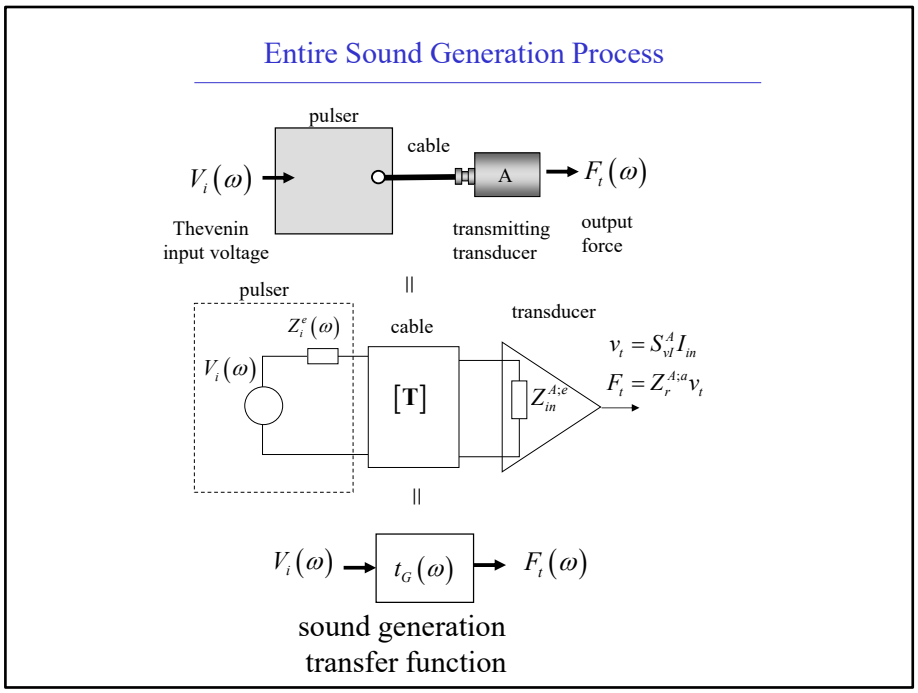
Sensitivity, Impedance



Thus, we can replace the transfer matrix model of the transducer by a model consisting of an electrical impedance and an ideal "converter" that is defined by the transducer sensitivity:

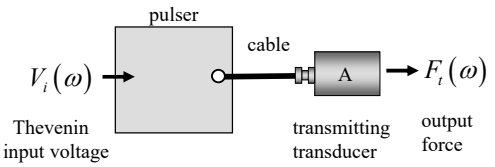


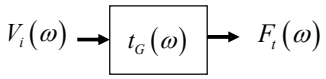
If we know the electrical impedance and sensitivity we can bypass knowing the 2x2 transfer matrix components and simply and completely model the sending transducer as this impedance and sensitivity combination. Later, we will show that like the electrical impedance of the transducer we can easily measure its sensitivity with purely electrical measurements.



We now can combine our pulser, cabling, and sending transducer models together and combine all these components into a single sound generation transfer function.

Entire Sound Generation Process





sound generation
transfer function

$$t_G(\omega) = \frac{F_t(\omega)}{V_i(\omega)} = \frac{Z_r^{A;a} S_{vl}^A}{\left(Z_{in}^{A;e} T_{11} + T_{12}\right) + \left(Z_{in}^{A;e} T_{21} + T_{22}\right) Z_i^e}$$

cable transfer function component(s)

Here is the explicit expression for this sound generation transfer function in terms quantities that are all either known or obtainable with electrical measurements.

References

Ristic, V.M., **Principles of Acoustic Devices**, John Wiley, 1983

Kino. G.S., **Acoustic Waves - Devices, Imaging and Analog Signal Processing**, Prentice-Hall, 1987.

Auld, B.A., **Acoustic Fields and Waves in Solids, 2nd Ed., Vols. I and II**, Krieger Publishing Co. , 1990.

Sacshe, W., and N.N. Hsu," Ultrasonic transducers for materials testing and their characterization," in **Physical Acoustics**, **Vol. XIV**, Eds. W.P. Mason and R.N. Thurston, 277-406, 1979.

Greeenspan, M., "Piston radiator: some extension of the theory," J. Acoust. Soc. Am., 65, 608-621, 1979.

Some references that explicitly deal with transducers.

Wave Propagation Fundamentals -1
Propagation of Plane and Spherical Waves

Propagating waves are a key part of an ultrasonic NDE measurement system. Plane waves and spherical waves serve as fundamental building blocks for describing the waves seen in NDE tests so we will examine those waves here.

Learning Objectives

Plane waves in fluids and solids

Elastic wave potentials

Spherical waves in fluids and solids

We will examine both **plane waves** and **spherical waves** in both fluid and elastic media.

Plane Waves - Fluid

Waves in a fluid satisfy the wave equation

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0$$

$p = \tilde{f}\left(t - x/c\right)$ is an arbitrary plane wave"pulse" solution of the wave equation traveling in the + x direction.

The function $p = \tilde{F}(f)\exp\left[2\pi i f\left(x/c - t\right)\right]$ is a harmonic plane wave traveling in the + x direction.

They are simply related through the Fourier Transform:

amplitude phase

$$\tilde{f}\left(t - x/c\right) = \int_{-\infty}^{+\infty} \tilde{F}(f)\exp\left[2\pi i f\left(x/c - t\right)\right] df$$

We have seen before that the pressure waves in a fluid satisfy the wave equation. We also saw that we could write down the general solutions of plane waves in the fluid either as pulses in the time domain or as harmonic waves in the frequency domain and we can relate those two domains through the Fourier transform.

Plane Waves-Fluid

we saw previously we can write a plane harmonic wave in different forms:

$$F(f)\exp[2\pi if(x/c-t)]$$
$$=F(f)\exp[ik(x-ct)]$$
$$=F(\omega)\exp[i\omega(x/c-t)]$$

For the last form we can also write

$$f(t-x/c)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}F(\omega)\exp[i\omega(x/c-t)]d\omega$$

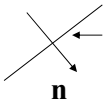
We also saw previously that we could write propagating plane harmonic waves in terms of different parameters. For solving wave problem we will generally use the forms shown above where we either use the frequency in cycles/sec or in rad/sec.

Plane Waves-Fluid

Summary: for a harmonic wave of $\exp(-i\omega t)$ time dependency, a plane wave traveling in the $\pm x$ direction is given by (dropping the common time term)

$$F\exp[\pm 2\pi i f x / c]$$
$$= F\exp[\pm i\omega x / c]$$
$$= F\exp[\pm ikx]$$

For a plane wave traveling in an arbitrary direction , **n**



$\mathbf{n}\cdot\mathbf{x}$ = constant plane

$$F\exp[ik\mathbf{n}\cdot\mathbf{x}]$$
$$= F\exp[i\mathbf{k}\cdot\mathbf{x}]$$

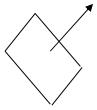
where $\mathbf{k} = k\,\mathbf{n}$ is called the wavenumber vector

Since all the harmonic waves have the same time dependency, which here is exp(-iωt), generally we will drop that common term and assume it is always implicitly present. All our plane wave results shown are for waves traveling in the plus or minus x-direction but we can easily write similar plane wave solutions for a wave traveling in an arbitrary direction, **n**, in three dimensions.

Plane Waves - Fluid and Solid Media

Waves in a fluid are governed by the scalar wave equation

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0$$



n (direction of propagation)

$$p = f(t - \mathbf{x}\cdot\mathbf{n}/c)$$

Waves in an isotropic elastic solid are governed by the vector Navier's equations

$$\mu\nabla^2\mathbf{u} + (\lambda+\mu)\nabla(\nabla\cdot\mathbf{u}) - \rho\frac{\partial^2\mathbf{u}}{\partial t^2} = 0$$

u … displacement vector
λ, μ … Lame constants
ρ … density

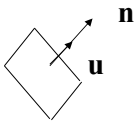
Here we show a general plane wave traveling in a fluid, where it is satisfies the scalar wave equation for the pressure. In an elastic solid, general waves are governed by **Navier's equations** for the displacements, not a wave equation directly. Navier's equations depend on the density, ρ , of the solid and two elastic Lame constants, λ, and μ.

Plane Waves-Solid

$$\mu\nabla^2\mathbf{u} + (\lambda+\mu)\nabla(\nabla\cdot\mathbf{u}) - \rho\frac{\partial^2\mathbf{u}}{\partial t^2} = 0$$

Navier's equations also have plane wave solutions. There are two types – P-waves and S-waves. These are "bulk" waves.

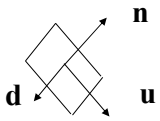
compressional (P) waves



$$\mathbf{u} = \mathbf{n}\, f\left(t - \mathbf{x}\cdot\mathbf{n}/c_p\right)$$

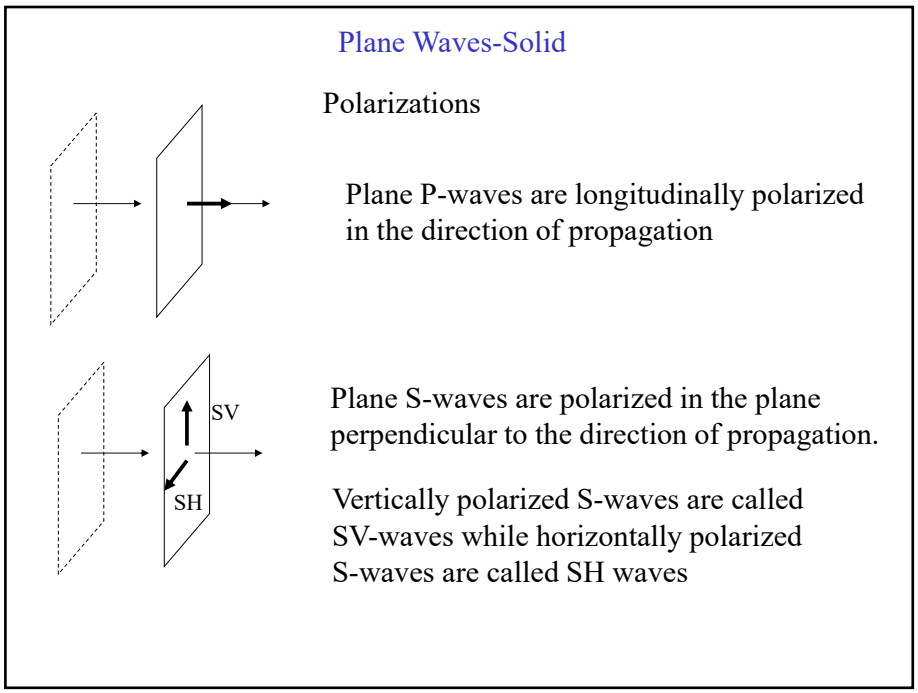
polarization

shear (S) waves



$$\mathbf{u} = \mathbf{n}\times\mathbf{d}\, g\left(t - \mathbf{x}\cdot\mathbf{n}/c_s\right)$$

Like a fluid, Navier's equations do have plane wave solutions. However, in a solid there are actually two types of solutions called **P-waves** and **S-waves**, which differ in their direction of displacement (called the **polarization**) and in their wave speed. The wave speed c_p is the P-wave speed and the wave speed c_s is the S- (shear) wave speed. Since both types of these waves travel in the bulk of a material, they are also called **bulk waves**. For most solids the P-wave speed is about twice the S- wave speed.



Plane P-waves are polarized in the direction of propagation while plane S-waves have a polarization in a plane perpendicular to the direction of propagation. If we take that polarization plane as a vertical plane then we can consider either vertically polarized S-waves, called **SV-waves**, or horizontally polarized S-waves called **SH-waves**.

Plane Waves-Solid

Bulk wave speeds

$$c_p = \sqrt{\frac{\lambda + 2\mu}{\rho}} = \sqrt{\frac{E(1-\nu)}{(1+\nu)(1-2\nu)\rho}}$$

E … Young's modulus
ν … Poisson's ratio
G = μ .. Shear Modulus

$$c_s = \sqrt{\frac{\mu}{\rho}} = \sqrt{\frac{G}{\rho}} = \sqrt{\frac{E}{2(1+\nu)\rho}} \qquad G = \frac{E}{2(1+\nu)}$$

Compressional wave speed in a bar $c_{bar} = \sqrt{\frac{E}{\rho}}$

Compressional wave speed in a plate $c_{plate} = \sqrt{\frac{E}{(1-\nu^2)\rho}}$

Bulk P- and S-waves travel with different wave speeds which can be written either in terms of the Lame constants of the elastic solid or Young's modulus, **E**, and Poisson's ratio , **v**. For S-waves we can also write the shear wave in terns of the shear modulus, **G**, which is related to E and v.

We should note that if we generate waves traveling in thin geometries like bars or plates, the P-waves travel with wave speeds different from the bulk wave speeds while the S-wave wave speed is unaffected. Shown are the P wave speeds for a bar and plate.

Plane Waves-Solid

$$\mu\nabla^2\mathbf{u} + (\lambda+\mu)\nabla(\nabla\cdot\mathbf{u}) - \rho\frac{\partial^2\mathbf{u}}{\partial t^2} = 0$$

Although Navier's equations are not wave equations, solutions to Navier's equations can be found in terms of potential functions that do satisfy wave equations

$$\mathbf{u} = \nabla\phi + \nabla\times\boldsymbol{\psi}$$

ϕ … scalar potential
$\boldsymbol{\psi}$ … vector potential

$$\nabla^2\phi - \frac{1}{c_p^2}\frac{\partial^2\phi}{\partial t^2} = 0 \qquad \text{P-waves}$$

$$\nabla^2\boldsymbol{\psi} - \frac{1}{c_s^2}\frac{\partial^2\boldsymbol{\psi}}{\partial t^2} = 0 \qquad \text{S-waves}$$

In Navier's equations we do not see directly wave equations. However, if we use the Helmholtz decomposition and represent the displacement in terms of potentials, then we do see separate wave equations for the P-waves and S-waves. Thus, solutions of elastic wave problems are often obtained in terms of these potentials.

Note that P-waves are also called compressional waves, longitudinal (L-waves), pressure waves, primary waves, dilatational waves, or irrotational waves while S-waves are also called shear waves, tangential or transverse (T-waves), secondary waves, equivoluminal waves, or rotational waves.

Material	c_p Compressional (P-wave) wave speed (m/s x 10^3)	c_s Shear (S-wave) wave speed (m/s x 10^3)	ρ Density ρ (kgm/m^3 x 10^3)	ρc_p Impedance (P-wave) (kgm/(m^2-s) x 10^6)
Air	0.33	--	0.0012	0.0004
Aluminum	6.42	3.04	2.70	17.33
Brass	4.70	2.10	8.64	40.6
Copper	5.01	2.27	8.93	44.6
Glass	5.64	3.28	2.24	13.1
Lucite	2.70	1.10	1.15	3.1
Nickel	5.60	3.00	8.84	49.5
Steel, mild	5.90	3.20	7.90	46.0
Titanium	6.10	3.10	4.48	27.3
Tungsten	5.20	2.90	19.40	101.0
Water	1.48	--	1.00	1.48

Table D.1 Acoustical properties of some common materials.

In the S.I. system 1 Rayl = 1kg/m^2-s so these values are in units of 10^6 Rayl or MRayl

Here is a short table illustrating the compressional and shear wave speeds, densities, and the specific acoustic impedances for P-waves of some common materials. Generally we see that compressional waves travel about twice as fast as shear waves. Also note that specific acoustic impedances are often given in units of megaRayls (MRayl) which units are defined above.

Plane Waves-Solid

Elastic Wave Potentials

$$\mathbf{u} = \nabla\phi + \nabla \times \boldsymbol{\psi}$$

For two-dimensional waves

$$\phi = \phi(x, y, t)$$

$$\psi_z = \psi(x, y, t),\ \psi_x = \psi_y = 0$$

displacements

$$u_x = \frac{\partial \phi}{\partial x} + \frac{\partial \psi}{\partial y}$$

$$u_y = \frac{\partial \phi}{\partial y} - \frac{\partial \psi}{\partial x}$$

$$u_z = 0$$

$$\kappa = \frac{c_p}{c_s}$$

stresses

$$\tau_{xx} = \mu\left[\kappa^2\nabla^2\phi + 2\left(\frac{\partial^2\psi}{\partial x\partial y} - \frac{\partial^2\phi}{\partial y^2}\right)\right]$$

$$\tau_{yy} = \mu\left[\kappa^2\nabla^2\phi - 2\left(\frac{\partial^2\psi}{\partial x\partial y} + \frac{\partial^2\phi}{\partial x^2}\right)\right]$$

$$\tau_{xy} = \mu\left[2\frac{\partial^2\phi}{\partial x\partial y} + \frac{\partial^2\psi}{\partial y^2} - \frac{\partial^2\psi}{\partial x^2}\right]$$

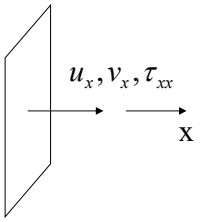
$$\tau_{zz} = \nu\left[\tau_{xx} + \tau_{yy}\right],\ \tau_{xz} = \tau_{yz} = 0$$

Potentials are often used to solve wave equations in elastic solids analytically. For 2-D problems one only needs to use scalar P-wave and S-wave potentials and those potentials (which are not physical quantities) can be related to the displacements and stresses as shown.

Plane Waves-Solid

Plane wave solutions

P-waves $\quad k_p = \omega / c_p$



$\phi = \Phi \exp\left(ik_p x - i\omega t\right)$ potential

$u_x = U_x \exp\left(ik_p x - i\omega t\right)$ displacement

$v_x = V_x \exp\left(ik_p x - i\omega t\right)$ velocity

$\tau_{xx} = T_{xx} \exp\left(ik_p x - i\omega t\right)$ normal stress

$U_x = ik_p \Phi, V_x = -i\omega U_x,$

$T_{xx} = -\rho c_p V_x$

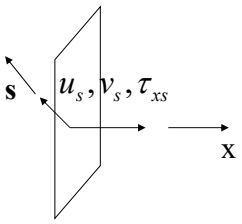
In an elastic solid we can consider plane P-wave solutions for the potential or we can consider the corresponding solution in terms of the displacement, the velocity, or the normal stress. All these waves have the same form where the amplitudes are related to each other as shown. Note that the stress and velocity amplitude relationship is like that of the pressure and velocity relationship except that there is a minus sign present because normal stress is defined to be positive when it is a tensile stress.

Plane Waves-Solid

S-waves

$k_s = \omega / c_s$

$\mathbf{s} = \mathbf{e}_x \times \mathbf{t}$



$\boldsymbol{\psi} = \Psi \mathbf{t} \exp\left(ik_s x - i\omega t\right)$ potential

$\mathbf{u} = U_s \mathbf{s} \exp\left(ik_s x - i\omega t\right)$ displacement

$\mathbf{v} = V_s \mathbf{s} \exp\left(ik_s x - i\omega t\right)$ velocity

$\tau_{xs} = T_{xs} \exp\left(ik_s x - i\omega t\right)$ shear stress

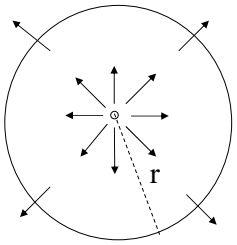
$U_s = ik_s \Psi, V_s = -i\omega U_s,$

$T_{xs} = -\rho c_s V_s$

In an elastic solid we can consider plane S-wave solutions for the potential or we can consider the corresponding solution in terms of the displacement, the velocity, or the shear stress. All these waves have the same form where the amplitudes are related to each other as shown. Note that in the shear wave case the polarization vector of the potential and the polarization vector of the displacement (or velocity) are different but they both lie in the plane of the wavefront and can be related to each other.

Spherical Waves-Fluid

Spherical waves arise physically from "point" sources. If a point source emits waves uniformly in all directions, then we expect the waves to depend only on the radial distance, r , from the point source.



Plane waves are important building blocks for modeling the propagation of waves. We will see that **spherical waves** are likewise a key type of building block. Physically, a spherical wave in a fluid can be though of as the waves arising from a point source that emits waves uniformly in all directions so that the waves depend only on the radial distance from the source.

Spherical Waves-Fluid

Consider spherical harmonic waves of exp(-iωt) time dependency. The equation of motion of the fluid and the wave equation are given by

$$-\nabla p = -i\omega\rho\mathbf{v}$$

$$\nabla^2 p + \frac{\omega^2}{c^2} p = 0$$

In terms of the distance, *r*, from the source, in spherical coordinates these equations are

$$\frac{\partial p}{\partial r} = i\omega\rho v_r$$

$$\frac{\partial^2 p}{\partial r^2} + \frac{2}{r}\frac{\partial p}{\partial r} + \frac{\omega^2}{c^2} p = 0$$

For harmonic spherical waves in a fluid Newton's law and the wave equation reduce to the forms shown.

Spherical Waves-Fluid

There are two solutions of the wave equation of the form

$$p = \frac{A}{r}\exp(ikr) + \frac{B}{r}\exp(-ikr)$$

$$k = \omega / c$$

The first term represents a wave traveling in the outward radial direction while the second term represents a wave converging on the source. Since the source only generates outward going waves, we must set $B = 0$. The pressure and radial velocity in the out going wave are then

$$p = \frac{A}{r}\exp(ikr)$$

$$v_r = \frac{A}{\rho c}\left[1 - \frac{1}{ikr}\right]\frac{\exp(ikr)}{r}$$

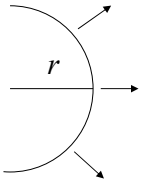
We can obtain solutions to the wave equation representing either outgoing waves from a point source or spherical waves that converge to the point. For the outgoing waves we have the forms shown for the pressure and velocity.

Spherical Waves-Fluid

In many cases we are only interested in the waves many wavelengths from the source, in which case kr >> 1 and we can write approximately

$$p = A\frac{\exp(ikr)}{r}$$

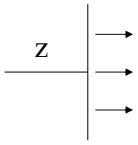
$$v_r = \frac{A}{\rho c}\frac{\exp(ikr)}{r}$$



Compare this to a plane wave traveling in the +z direction where

$$p = A\exp(ikz)$$

$$v_z = \frac{A}{\rho c}\exp(ikz)$$



If we are primarily interested in the waves many wavelengths from the source, then the factor kr >>1 and the pressure and velocity take the simpler forms shown above. These look very similar to the plane wave case except for the 1/r spreading factor which is absent in the plane wave case and causes the amplitude of the spherical wave to become smaller as the distance from the source increases.

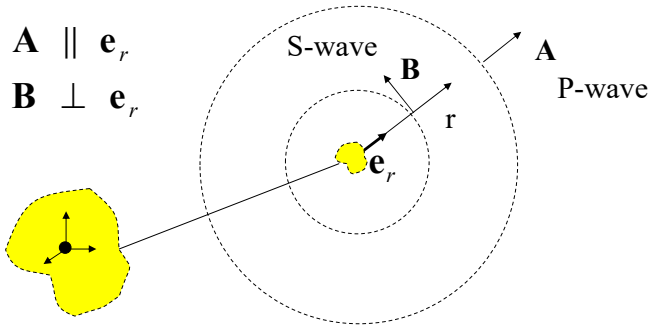
Spherical Waves-Elastic Solid

Spherical waves from point sources in an elastic solid have a more complex structure in general, but for kr >>1 they are similar to that of a fluid:

$$\mathbf{u} = \mathbf{A}\frac{\exp(ik_p r)}{r} + \mathbf{B}\frac{\exp(ik_s r)}{r}$$

$\mathbf{A} \parallel \mathbf{e}_r$

$\mathbf{B} \perp \mathbf{e}_r$



In elastic solids, P- and S- waves from a point source have a more complex behavior but for kr >>1 the displacement in the waves has a radial dependency that looks much like that in a fluid. The amplitudes are now vector quantities that are functions of the angles present in a spherical coordinate system and have polarization vectors that are either in the radial direction, **A**, of propagation (for P-waves) or in a direction, **B**, orthogonal to that radial direction (for S-waves).

Wave Propagation Fundamentals -2
Reflection and Transmission

NDE immersion tests involve multiple media (i.e. a fluid and solid) so it is important to understand how waves reflect and transmit at the interfaces between different media.

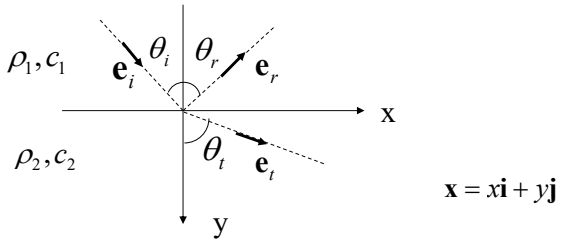
Learning Objectives

Reflection and transmission coefficients
- Snell's law
- critical angles
- acoustic impedance
- amplitudes/intensities
- inhomogeneous waves
- Stokes' relations

Fluid-fluid and fluid-solid examples

In this section we will see how plane waves reflect and transmit at planar interfaces between fluids or between fluids and elastic solids. It will be shown that there are a number of important concepts that govern reflection/transmission behavior including **Snell's law**, **critical angles**, and **Stokes relations**. In some cases we will give detailed derivation of results but in other cases we will simply state important results without proof.

Fluid-Fluid Problem

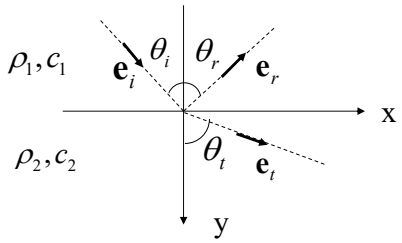


$\mathbf{x} = x\mathbf{i} + y\mathbf{j}$

pressure in the:

incident wave	$P_i \exp\left[ik_1 \mathbf{e}_i \cdot \mathbf{x}\right]$	$\mathbf{e}_i = \sin\theta_i \mathbf{i} + \cos\theta_i \mathbf{j}$
reflected wave	$P_r \exp\left[ik_1 \mathbf{e}_r \cdot \mathbf{x}\right]$	$\mathbf{e}_r = \sin\theta_r \mathbf{i} - \cos\theta_r \mathbf{j}$
transmitted wave	$P_t \exp\left[ik_2 \mathbf{e}_t \cdot \mathbf{x}\right]$	$\mathbf{e}_t = \sin\theta_t \mathbf{i} + \cos\theta_t \mathbf{j}$

Consider the problem of the reflection of a plane pressure wave (traveling at an angle as shown) from a plane interface between two different fluids. This case is normally not encountered in practice but it will be a simple case that captures much of the physics involved in more practical cases. An incident harmonic wave will generate reflected and transmitted plane waves whose solutions we can write as indicated above., where we will drop the common exp(-iωt) terms.



$$p_1 = P_i \exp\left[ik_1\left(x\sin\theta_i + y\cos\theta_i\right)\right] + P_r \exp\left[ik_1\left(x\sin\theta_r - y\cos\theta_r\right)\right]$$

$$p_2 = P_t \exp\left[ik_2\left(x\sin\theta_t + y\cos\theta_t\right)\right]$$

from the equation of motion $-\dfrac{\partial p}{\partial y} = -i\omega\rho\, v_y$, so

$$\left(v_y\right)_1 = \frac{P_i\cos\theta_i}{\rho_1 c_1}\exp\left[ik_1\left(x\sin\theta_i + y\cos\theta_i\right)\right] - \frac{P_r\cos\theta_r}{\rho_1 c_1}\exp\left[ik_1\left(x\sin\theta_i - y\cos\theta_i\right)\right]$$

$$\left(v_y\right)_2 = \frac{P_t\cos\theta_t}{\rho_2 c_2}\exp\left[ik_2\left(x\sin\theta_t + y\cos\theta_t\right)\right]$$

In the first medium the incident and reflected waves are present while in the second medium we only have a transmitted waves. From Newton's law we can relate the derivative of the pressure in the y-direction to the velocity component in that direction, so taking the derivatives of the pressure expressions in the two media we can obtain the y-velocity expressions for those media.

Boundary conditions: on y = 0 $p_1 = p_2$

$$\left(v_y\right)_1 = \left(v_y\right)_2$$

$$P_i \exp\left(ik_1 x \sin\theta_i\right) + P_r \exp\left(ik_1 x \sin\theta_r\right) = P_t \exp\left(ik_2 x \sin\theta_t\right)$$

$$\frac{P_i \cos\theta_i}{\rho_1 c_1} \exp\left(ik_1 x \sin\theta_i\right) - \frac{P_r \cos\theta_r}{\rho_1 c_1} \exp\left(ik_1 x \sin\theta_r\right) = \frac{P_t \cos\theta_t}{\rho_2 c_2} \exp\left(ik_2 x \sin\theta_t\right)$$

Phase matching: $k_1 \sin\theta_i = k_1 \sin\theta_r = k_2 \sin\theta_t$

$\Rightarrow$ $\theta_i = \theta_r$

angle of incidence
= angle of reflection

$$\frac{\sin\theta_i}{c_1} = \frac{\sin\theta_t}{c_2}$$

Snell's law

At the interface (y = 0) the pressure and the y-component of the velocity must be continuous (but the x-component of the velocity need not be continuous because we are modeling the fluids as ideal (i.e. non-viscous) fluids). Placing our pressure and velocity expressions into these boundary conditions we obtain the two equations shown. Since we must have these equations satisfied for all x we must have their phase terms all match which leads to two results: **the angle of incidence must be equal to the angle of reflection**, and the incident and transmitted wave angles must satisfy **Snell's law**.

$$P_i + P_r = P_t$$

$$\frac{P_i \cos\theta_i}{\rho_1 c_1} - \frac{P_r \cos\theta_i}{\rho_1 c_1} = \frac{P_t \cos\theta_t}{\rho_2 c_2} \qquad \cos\theta_t = \sqrt{1 - \frac{c_2^2}{c_1^2}\sin^2\theta_i}$$

Solving, we find the transmission and reflection coefficients (pressure ratios)

$$T_p = \frac{P_t}{P_i} = \frac{2\rho_2 c_2 \cos\theta_i}{\rho_1 c_1 \cos\theta_t + \rho_2 c_2 \cos\theta_i}$$

$$R_p = \frac{P_r}{P_i} = \frac{\rho_2 c_2 \cos\theta_i - \rho_1 c_1 \cos\theta_t}{\rho_1 c_1 \cos\theta_t + \rho_2 c_2 \cos\theta_i}$$

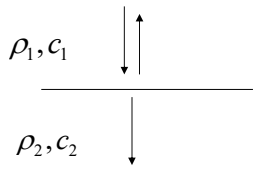
or, in terms of velocity ratios $(P = \rho c V)$

$$T_v = \frac{V_t}{V_i} = \frac{2\rho_1 c_1 \cos\theta_i}{\rho_1 c_1 \cos\theta_t + \rho_2 c_2 \cos\theta_i}$$

$$R_v = \frac{V_r}{V_i} = \frac{\rho_2 c_2 \cos\theta_i - \rho_1 c_1 \cos\theta_t}{\rho_1 c_1 \cos\theta_t + \rho_2 c_2 \cos\theta_i}$$

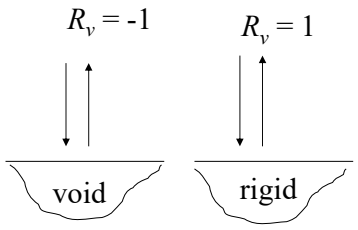
Eliminating the common phase terms gives us two equations for the reflected and transmitted wave amplitudes which we can solve to determine the transmission and reflection transmission components (based on pressure amplitude ratios). Using the plane wave relationship between pressure and velocity amplitudes we can alternately express those transmission and reflection coefficients in terms of velocity ratios.

Special Case - normal incidence



$$T_p = T_v = \frac{2\rho_1 c_1}{\rho_1 c_1 + \rho_2 c_2} = \frac{2z_1^a}{z_1^a + z_2^a}$$

$$R_p = R_v = \frac{\rho_2 c_2 - \rho_1 c_1}{\rho_1 c_1 + \rho_2 c_2} = \frac{z_2^a - z_1^a}{z_1^a + z_2^a}$$



$z^a = \rho c$ … plane wave specific impedance

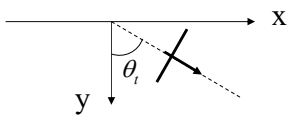
In the special case where the incident plane wave strikes the plane interface at normal incidence the velocity based transmission and reflection coefficients are of a much simpler form. We see that these coefficients are controlled by the plane wave specific acoustic impedances.

If the second medium is a void (a free surface), then the reflection coefficient is just -1, showing the incident wave is totally reflected with a change in sign. However, the velocity polarization of the reflected wave was taken in this case to be in the –y direction so that the y-velocity at the free surface is actually double that of the incident wave. In contrast, the total pressure due to the incident and reflected wave amplitudes is zero, as it should be since we have a free surface.

If the second medium is taken to be an infinitely rigid medium, then the reflection coefficient is +1 and so the total y-velocity at the rigid interface is just zero. The pressures of the incident and reflected waves, however, add so that the total pressure at the interface is just double that of the incident wave.

Critical angle $\theta_{cr} = \sin^{-1}(c_1 / c_2)$

Let $c_2 > c_1$. Then for $\theta_i < \sin^{-1}\left(\frac{c_1}{c_2}\right)$ $\quad \cos\theta_t = \sqrt{1 - \frac{c_2^2}{c_1^2}\sin^2\theta_i}$

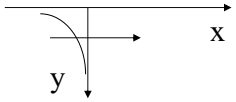


$$p_2 = P_t \exp\left[ik_2\left(x\sin\theta_t + y\cos\theta_t\right)\right]$$

transmitted plane wave

Whereas for $\theta_i > \sin^{-1}\left(\frac{c_1}{c_2}\right)$ $\quad \cos\theta_t = i\,\mathrm{sgn}\,\omega\sqrt{\frac{c_2^2}{c_1^2}\sin^2\theta_i - 1}$

$$\mathrm{sgn}\,\omega = \begin{cases} +1 & \omega > 0 \\ -1 & \omega < 0 \end{cases}$$



$$p_2 = P_t \exp\left(\frac{-|\omega| y}{c_2}\sqrt{\frac{c_2^2}{c_1^2}\sin^2\theta_i - 1}\right)\exp\left(ik_1 x\sin\theta_i\right)$$

inhomogeneous wave

If the wave speed of the second medium is higher than the wave speed in the first medium (containing the incident wave) then we will see that there is **critical angle,** $\sin^{-1}(c_1/c_2)$, which controls the behavior of the transmitted wave. For incident angles less than the critical angle the expression for the cosine of the transmitted angle (which involves a square root) is real so that we just see an ordinary plane wave expression for the transmitted wave. However, if the incident angle exceeds the critical angle then the cosine of the transmitted wave becomes imaginary. Depending on whether the frequency is positive or negative, we must change the sign on this cosine term so that the pressure never grows exponentially large in the y-coordinate, which would give an unphysical behavior. Thus, in this case we get a wave traveling parallel to the interface with an exponentially decreasing amplitude away from the boundary. This is called an **inhomogeneous wave**.

For there to be a critical angle the second medium must be the faster medium since there is no critical angle where $\sin^{-1}(c_1/c_2)$ is real if the second medium is slower than the first.

Relationship between amplitudes and intensities

Intensity, I , = time average power flux in the wave

Fluid:

$$I = \frac{P^2}{2\rho c} = \frac{\rho c V^2}{2}$$

P … pressure amplitude
V … velocity amplitude

Solid:

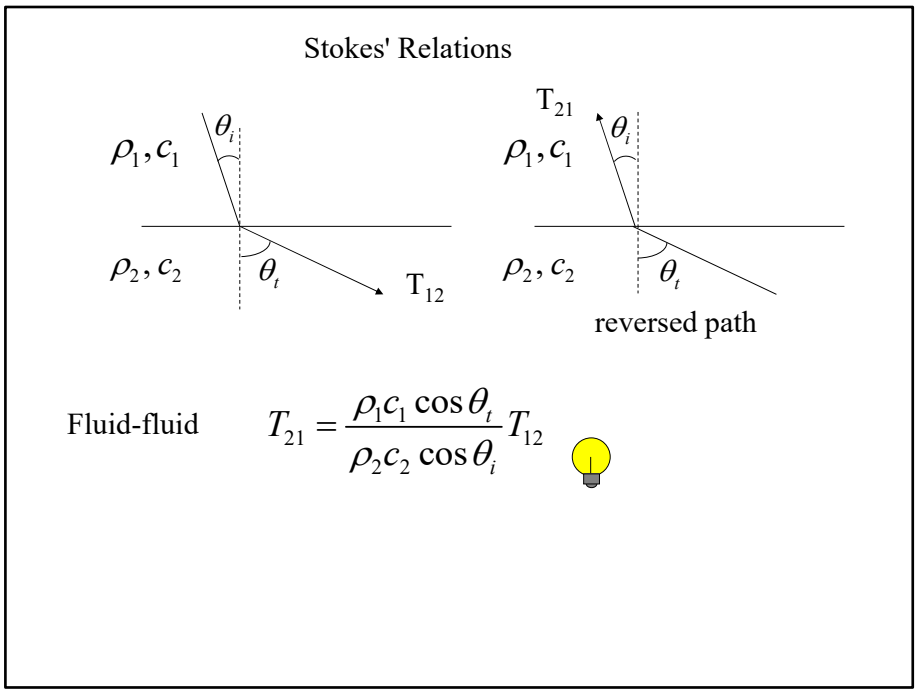
P-wave $$I = \frac{T_{nn}^2}{2\rho c_p} = \frac{\rho c_p V_n^2}{2}$$

T_{nn} … normal stress amplitude
T_{ns} … shear stress amplitude

S-wave $$I = \frac{T_{ns}^2}{2\rho c_s} = \frac{\rho c_s V_s^2}{2}$$

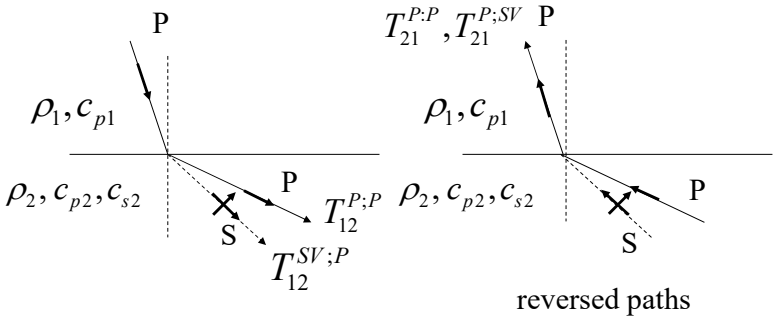
V_n .. normal velocity amplitude
V_s … tangential velocity amplitude

In some cases we want to know the energy being carried in a harmonic wave. We can represent such energy through an **intensity**, I, which defines a time average power flux in the wave. Shown are expressions for the intensities of a plane wave travelling in a fluid or a solid in terms of the underlying amplitudes of pressure, velocity, or stress.



In a pulse-echo inspection the same transducer is used as both a transmitter and receiver of ultrasound. In that case the waves that are scattered from a flaw back to the transducer follow a completely reversed path from the incident waves. and there can be transmission coefficients involved in both directions. One can show that in the case of two fluids these transmission coefficients are related through the **Stokes relations** given above.

Stokes' Relations

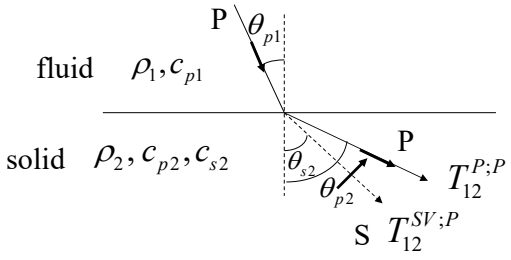


Fluid-solid $$T_{21}^{P;P} = \frac{\rho_1 c_{p1} \cos\theta_{p2}}{\rho_2 c_{p2} \cos\theta_{p1}} T_{12}^{P;P}$$

$$T_{21}^{P;SV} = \frac{\rho_1 c_{p1} \cos\theta_{s2}}{\rho_2 c_{s2} \cos\theta_{p1}} T_{12}^{SV;P}$$

If we consider a fluid-solid interface, since there can be both P-waves and S-waves in the solid we have the two Stokes relations given above. Note that we are taking the incident wave to be propagating in a vertical plane here. In this case we can show that the S-wave will have its polarization also in that plane and we will call it an SV-wave.

Transmission Coefficients (Fluid-Solid interface)
based on velocity ratios



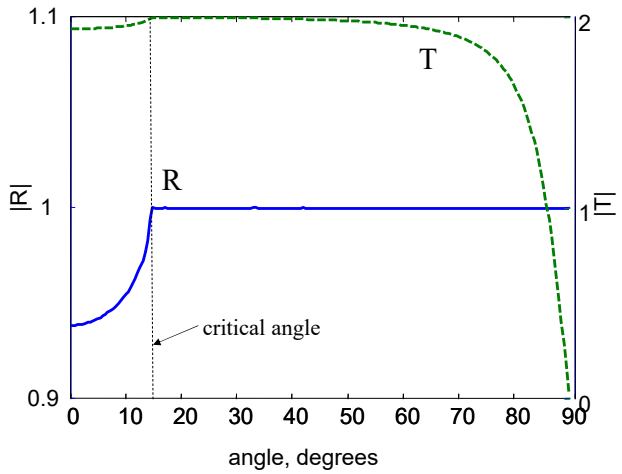
$$T_{12}^{P;P} = \frac{2\cos\theta_{p1}\left[1-2\left(\sin\theta_{s2}\right)^2\right]}{\cos\theta_{p2} + \frac{\rho_2 c_{p2}}{\rho_1 c_{p1}}\cos\theta_{p1}\left[4\left(\frac{c_{s2}}{c_{p2}}\right)^2 \sin\theta_{s2}\cos\theta_{s2}\sin\theta_{p2}\cos\theta_{p2} + 1 - 4\left(\sin\theta_{s2}\cos\theta_{s2}\right)^2\right]}$$

$$T_{12}^{SV;P} = \frac{-4\cos\theta_{p1}\cos\theta_{p2}\sin\theta_{s2}}{\cos\theta_{p2} + \frac{\rho_2 c_{p2}}{\rho_1 c_{p1}}\cos\theta_{p1}\left[4\left(\frac{c_{s2}}{c_{p2}}\right)^2 \sin\theta_{s2}\cos\theta_{s2}\sin\theta_{p2}\cos\theta_{p2} + 1 - 4\left(\sin\theta_{s2}\cos\theta_{s2}\right)^2\right]}$$

We can solve for the plane wave reflection and transmission coefficients at a fluid-solid interface in the same manner as done for the fluid-fluid interface but with more algebra. In an immersion inspection the waves are generated in a fluid, so we are primarily interested in the transmission coefficients into the solid component being inspected. Shown are those **transmission coefficients based on velocity ratios**.

```
>> ang = linspace(0, 89.9, 200);
>> [R,T] = pressure_coeffs(1.0, 7.9, 1480, 5900, ang);
>> plotyy(ang, abs(R), ang, abs(T), @myplot1, @myplot2)
>> xlabel('angle, degrees')
>> ylabel('|R|')
```

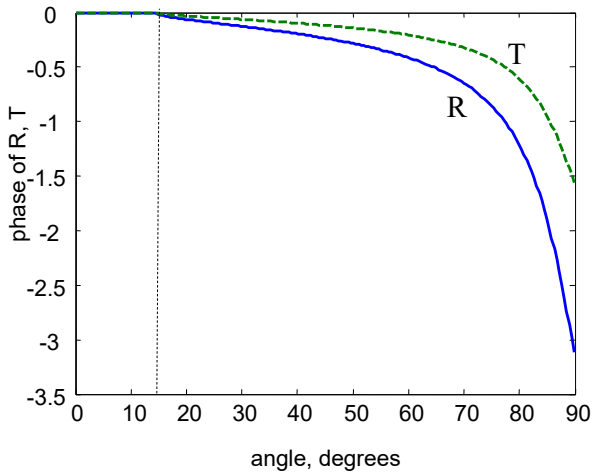
Fluid-Fluid R, T coefficients (magnitude) water-steel pressure-based



Here are some MATLAB results showing the refection and transmission coefficients obtained for a fluid-fluid interface where the first medium was water and the second fluid P-wave speed was taken to be the same as the compressional wave speed of steel. One can see a critical angle here at about 15 degrees.

```
>> plot(ang, unwrap(angle(R)), ang, unwrap(angle(T)), '--')
>> xlabel('angle, degrees')
>> ylabel('phase of R, T')
```

Fluid-Fluid R, T coefficients (phase) water-steel pressure-based

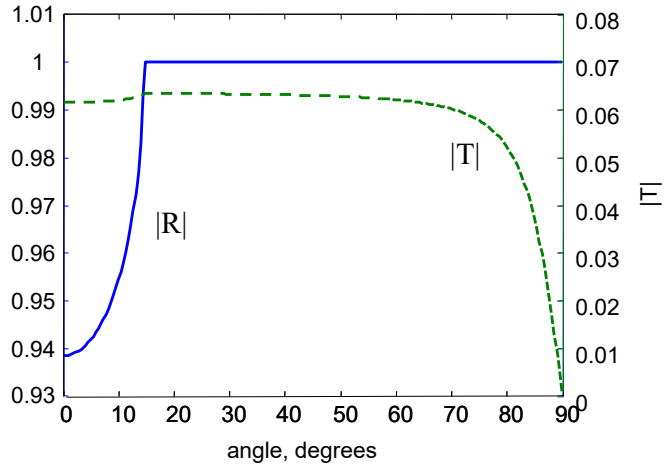


Here is the same fluid-fluid model results for the phase of the reflection and transmission coefficients. Note that below the critical angle the coefficients are real so the phase angles are zero.

```
>> [R,T] = velocity_coeffs(1.0, 7.9, 1480, 5900, ang);
>> plotyy(ang, abs(R), ang, abs(T), @myplot1, @myplot2)
>> xlabel('angle, degrees')
```

Fluid-Fluid R, T coefficients (magnitude) water-steel velocity-based

R is same as for pressure, T is much smaller

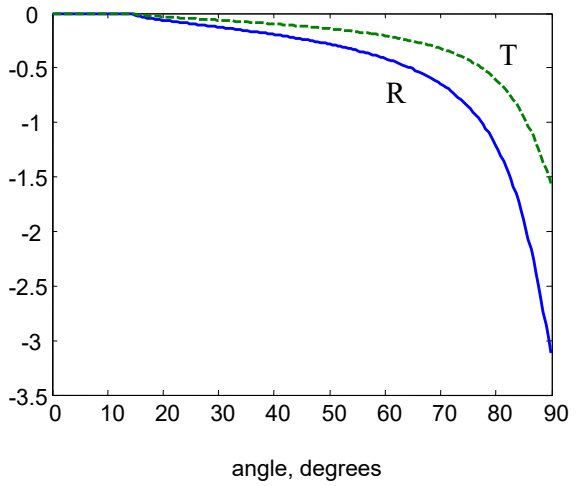


If we use reflection and transmission coefficients based on velocity ratios, then the reflection coefficient stays the same but the transmission coefficient is much smaller.

```
>> [R,T] = velocity_coeffs(1, 7.9, 1480, 5900, ang);
>> plot(ang, unwrap(angle(R)), ang, unwrap(angle(T)), '--')
>> xlabel('angle, degrees')
```

Fluid-Fluid R, T coefficients (phase) water-steel velocity-based

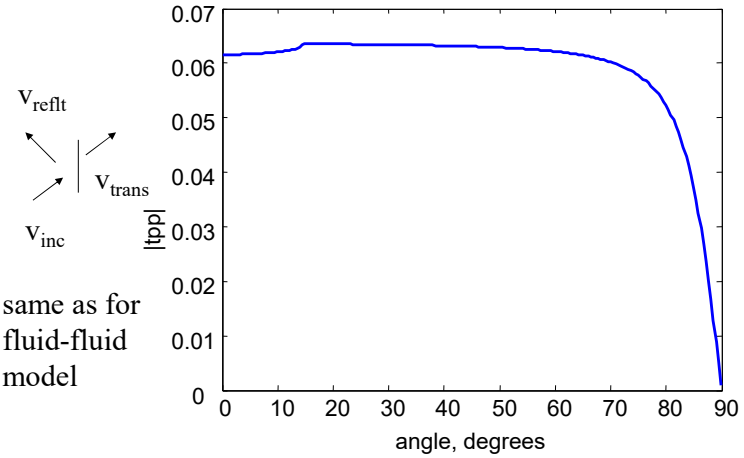
phases are same as for pressure



The phases for the velocity ratios are the same as for the pressure ratios.

```
>> [tpp,tss] = trans2(ang,1.0, 7.9,1480,5900,0);
>> plot(ang, abs(tpp))
>> xlabel('angle, degrees')
>> ylabel(' |tpp|')
```

Fluid-Solid Tpp coefficient (magnitude) water-steel velocity-based with c_s =0



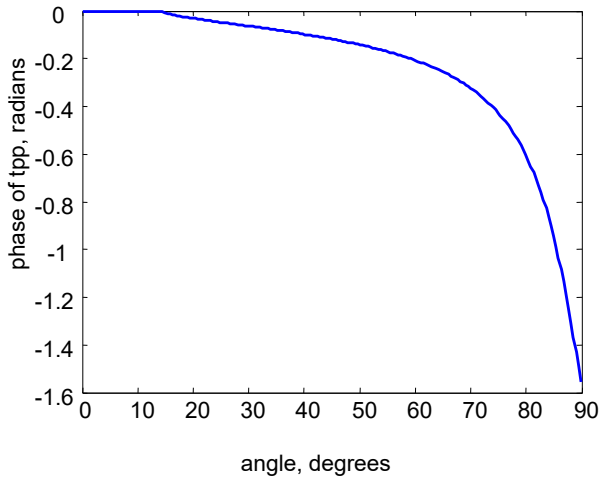
same as for fluid-fluid model

Here is the transmission coefficient for P-waves using the full solid model for a water-steel interface but where we set the shear wave speed of the solid to be zero. In this case, we do recover the fluid-fluid result

```
>> plot(ang, unwrap(angle(tpp)))
>> xlabel('angle, degrees')
>> ylabel('phase of tpp, radians')
```

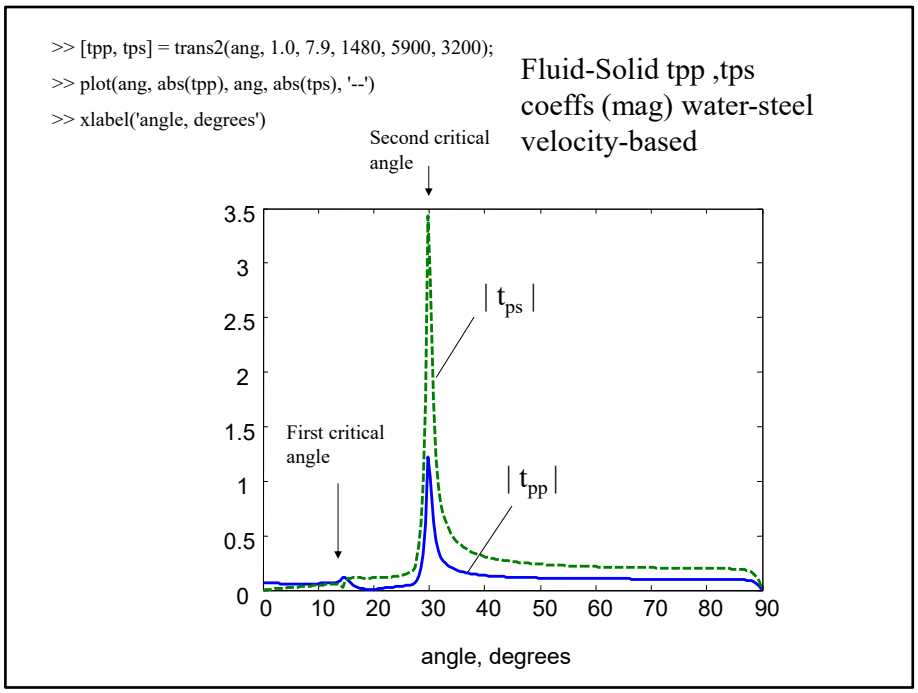
Fluid-Solid Tpp coefficient (phase) water-steel velocity-based with $c_s = 0$

same as for fluid-fluid model



The phase is also the same for the fluid-solid and fluid-fluid models in this water-steel case.

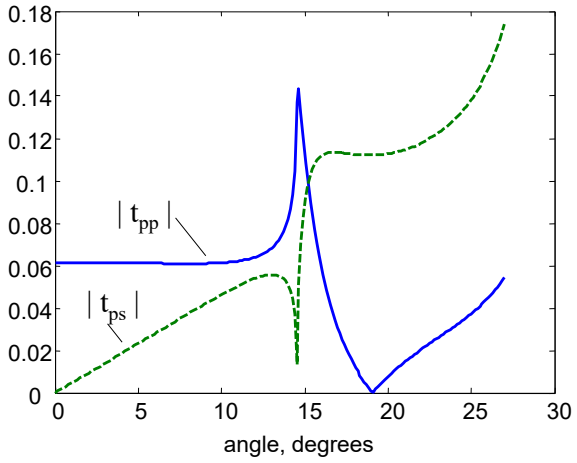
```
>> [tpp, tps] = trans2(ang, 1.0, 7.9, 1480, 5900, 3200);
>> plot(ang, abs(tpp), ang, abs(tps), '--')
>> xlabel('angle, degrees')
```



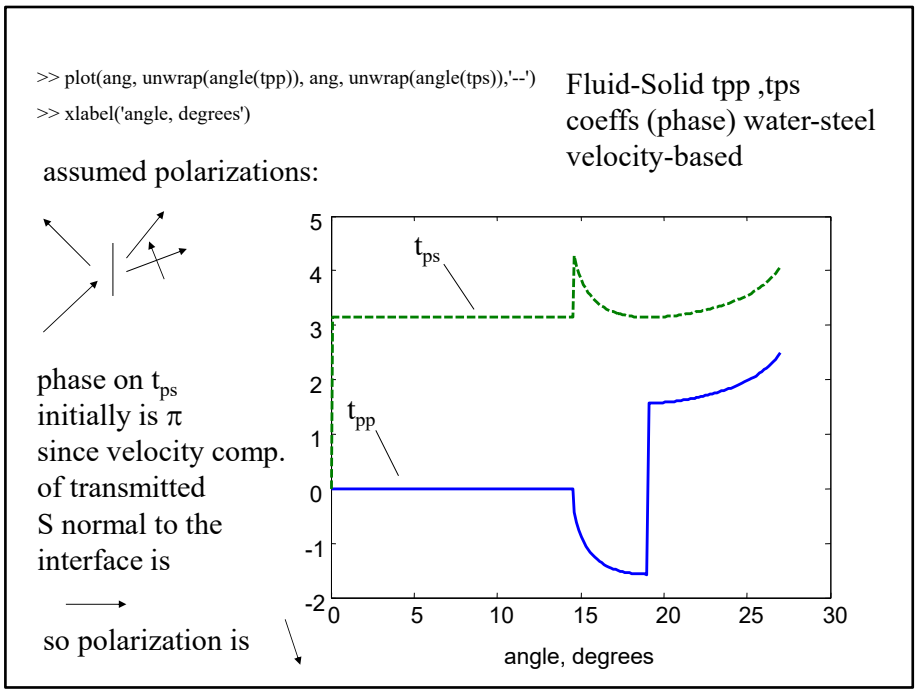
Here now are the actual fluid-solid results for both transmitted P-waves and S-waves, showing the magnitude of the transmission coefficients based on velocity ratios. We see that there are two angles where the behavior changes. These are critical angles which we will describe shortly.

```
>> ang =linspace(0, 27, 200);
>> [tpp, tps] = trans2(ang, 1.0, 7.9, 1480, 5900, 3200);
>> plot(ang, abs(tpp), ang, abs(tps),'--')
>> xlabel('angle, degrees')
```

Fluid-Solid tpp ,tps coeffs (mag) water-steel velocity-based (below second critical angle)



This is an expanded view of the transmission coefficients below the second critical angle. We see that the P-P coefficient is nearly a constant up to near the first critical angle while P-S coefficient has a linearly increasing behavior.



Here are the phases of the transmitted waves. Below the first critical angle the phase of the P-to-P coefficient is zero but the P-S coefficient has a phase of pi. This occurs because the actual polarization of the transmitted wave is opposite to that of the assumed polarization and recall exp(i π) =-1 so this phase of π just means the amplitude of motion in the wave is in a direction opposite to that we assumed.

Fluid-fluid R,T coefficients, pressure-based

```
function [R,T] =pressure_coeffs(d1, d2, c1, c2, ang)
iang = (ang.*pi)./180;
sint =(c2/c1)*sin(iang);
cost = sqrt(1-sint.^2).*(sint <= 1) + i.*sqrt(sint.^2 -1).*(sint > 1);
R = ((c2*d2).*cos(iang) - (c1*d1).*cost)./ ...
((c2*d2).*cos(iang) + (c1*d1).*cost);
T = ((2*d2*c2).*cos(iang))./((c2*d2).*cos(iang) + (c1*d1).*cost);
```

Fluid-fluid R, T coefficients, velocity-based

```
function [R,T] =velocity_coeffs(d1, d2, c1, c2, ang)
iang = (ang.*pi)./180;
sint =(c2/c1)*sin(iang);
cost = sqrt(1-sint.^2).*(sint <= 1) + i.*sqrt(sint.^2 -
1).*(sint > 1);
R = ((c2*d2).*cos(iang) - (c1*d1).*cost)./ ...
((c2*d2).*cos(iang) + (c1*d1).*cost);
T = ((2*d1*c1).*cos(iang))./((c2*d2).*cos(iang) +
(c1*d1).*cost);
```

Here are the MATLAB functions for the fluid-fluid models we have discussed.

Fluid-Solid P-P, P-S transmission coefficients

```
function [tpp,tps] = trans2(inc, d1, d2, cp1, cp2, cs2)
iang = (inc.*pi)./180;
sinp = (cp2/cp1)*sin(iang);
sins =(cs2/cp1)*sin(iang);
cosp= (i*sqrt(sinp.^2 - 1)).*(sinp >= 1) + (sqrt(1 - sinp.^2)).*(sinp < 1);
coss = (i*sqrt(sins.^2 - 1)).*(sins >= 1) + (sqrt(1 - sins.^2)).*(sins < 1);
denom = cosp + (d2/d1)*(cp2/cp1)*sqrt(1-sin(iang).^2).*(4.*((cs2/cp2)^2).*(sins.*coss.*sinp.*cosp) ...
                        + 1 - 4.*(sins.^2).*(coss.^2));
tpp = (2*sqrt(1 - sin(iang).^2).*(1 - 2*(sins.^2)))./denom;
tps = -(4*cosp.*sins.*sqrt(1 - sin(iang).^2))./denom;
```

Here is the MATLAB function for the fluid-solid model transmission coefficients.

Wave Propagation Fundamentals -3
Ultrasonic Attenuation

Waves propagating in real material will experience attenuation due to a number of sources. Our discussion of wave propagation so far has ignored attenuation so we will now look at some methods for including this attenuation in our models.

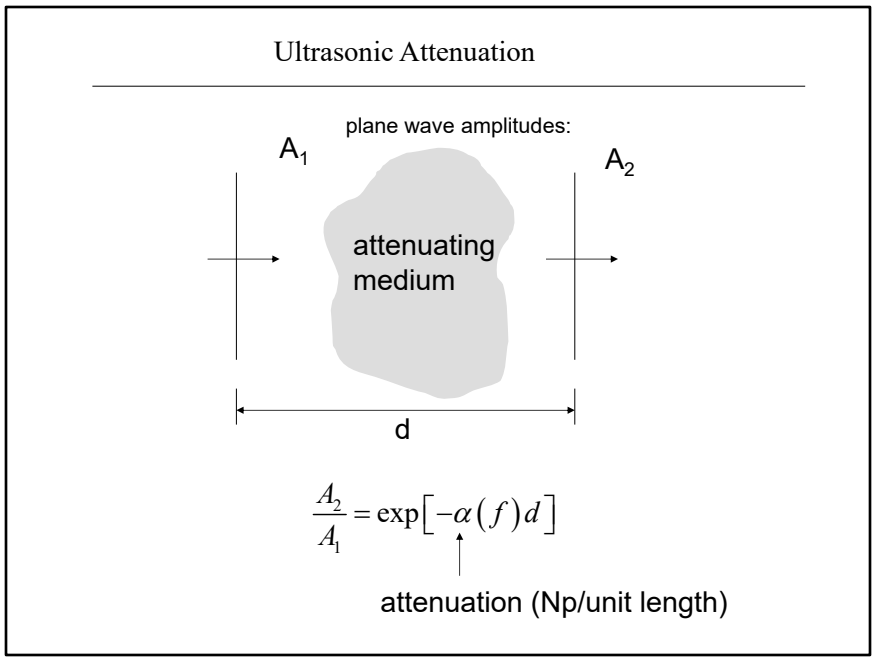
Learning Objectives

Attenuation definitions

Simple method for measuring average attenuation

Model-based approach for determining attenuation as a function of frequency

We will give a very simple method for obtaining an average attenuation value for a material and describe a more detailed approach that will allow us to include a frequency dependent attenuation in wave propagation and wave scattering calculations.



Attenuation in real materials is a complex process. Attenuation may be due, for example, as a result of the scattering of waves from the small grains that exist in a material, thus reducing the amplitude of a propagating wave. In our models of wave propagation we will not try to describe those attenuation processes in detail. Instead, we will use a simple plane wave model where will assume that the attenuation effects can be included by multiplying the wave amplitudes obtained for an ideal (on-attenuating) medium by an exponential decaying factor containing a frequency dependent attenuation coefficient and the distance of propagation. The attenuation coefficient is measured in Nepers/unit length where a Neper (Np) is a non-dimensional quantity.

John Napier of Merchistoun (1550 – 4 April 1617), nicknamed Marvellous Merchistoun, was a Scottish mathematician, physicist, astronomer/astrologer and 8th Laird of Merchistoun. He is most remembered as the inventor of logarithms and Napier's bones, and for popularizing the use of the decimal point. Napier's birth place, Merchiston Tower, Edinburgh, Scotland, is now part of Napier University. He is buried in St Cuthbert's Church, Edinburgh

Here is a short description of Napier. The Neper unit is derived from his name.

Example: water

$$\alpha_w(f) = 25.3\times10^{-15} f^2 \quad Np/m \qquad f \ldots Hz$$

Attenuation is affected by many variables so that generally it is not possible to give generic values that are useful for quantitative analyses.

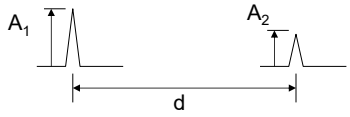
Thus, attenuation generally must be obtained experimentally for the material in question

When attenuation values are quoted they are often given as average values in decibels/unit length

$$\overline{\alpha}_{dB/l} = 8.686\overline{\alpha}_{Np/l}$$

We can easily measure the attenuation coefficient in water as a function of frequency. The result is shown above, where we see that the attenuation increases like the square of the frequency. Here the attenuation coefficient is given in Nepers/meter and the frequency, f, is measured in Hertz (Hz). For other materials it is not possible to give similar generic results since the attenuation is controlled by many aspects of how the material is manufactured and it is usually necessary to measure the attenuation on a sample of the material being used. Attenuation values are also often quoted in terms of decibels/ unit length but the conversion from Nepers/unit length to decibels/unit length is simple, as shown (see the next slide for how this conversion is obtained).

plane wave pulse traveling in an attenuative medium:



average attenuation (Np/l)

$$\frac{A_2}{A_1} = \exp\left[-\bar{\alpha}_{Np/l}\, d\right]$$

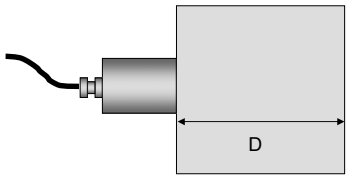
average attenuation (dB/l)

$$\begin{aligned}\bar{\alpha}_{dB/l} &= -\frac{20\log_{10}\left(\frac{A_2}{A_1}\right)}{d} \\ &= \left(20\log_{10} e\right)\bar{\alpha}_{Np/l} \\ &= 8.686\,\bar{\alpha}_{Np/l}\end{aligned}$$

One can measure the amplitude changes of a waveform over a distance and determine an average attenuation for the material. This attenuation can also be expressed in either Nepers/unit length or dB/unit length. We need a more detailed measure of the attenuation as a function of frequency to include in our ultrasonic models but a measurement of the average attenuation can give us a rough indication of the importance of attenuation in a material.

Simple method to determine average attenuation

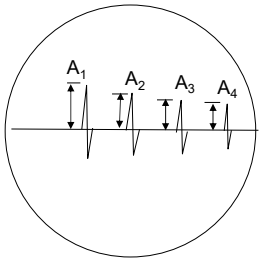
To model the effects of attenuation on the characteristics of a signal, we must be able to find the attenuation of a material as a function of frequency. However, in some applications we may be content to obtain an average attenuation value which describes the overall behavior of the amplitude of the signal. We can obtain such the average attenuation of a solid sample, for example, by using the pulse-echo setup shown.



On the oscilloscope screen we will see a series of evenly spaced pulses that are decreasing in amplitude. These are the waves that have been reflected from the back surface of the block one or more times.

If all we are interested in is the effects of attenuation on the signal amplitude, we can use a very simple pulse-echo measurement to obtain the average attenuation. We place a transducer on a block of the material and examine the waves that are reflected one or more times from the back surface.

Here is an example A-scan we might see:



Let the measured amplitudes of these signals be A_1, A_2, … etc. as shown

If the back surface of the block is in the far-field of the transducer, the incident and reflected waves will behave like attenuated spherical waves so we could express the voltage of these signals, $v(t)$, in the form:

$$v(t) = \frac{g(t-2D/c)}{2D}\exp\left(-2\bar{\alpha}_{Np/l}D\right) + \frac{g(t-4D/c)}{4D}\exp\left(-4\bar{\alpha}_{Np/l}D\right) + \frac{g(t-6D/c)}{6D}\exp\left(-6\bar{\alpha}_{Np/l}D\right) + \ldots$$

where the waveform shape is given by the g(t) function and $\bar{\alpha}_{Np/l}$ is the average attenuation of the block (in Np/l)

Here is an example response we might see. If the back surface of the block is sufficiently far from the transducer (i.e. in the “far field” of the transducer, which we define later) the reflected waves will look attenuated spherical waves so we can write the time domain response as shown.

If we let g_{max} be the maximum amplitude of the g function then the amplitudes shown are given by

$$A_1 = \frac{g_{\max}}{2D}\exp\left(-2\bar{\alpha}_{Np/l}D\right)$$

$$A_2 = \frac{g_{\max}}{4D}\exp\left(-4\bar{\alpha}_{Np/l}D\right)$$

$$A_3 = \frac{g_{\max}}{6D}\exp\left(-6\bar{\alpha}_{Np/l}D\right)$$

$$A_4 = \frac{g_{\max}}{8D}\exp\left(-8\bar{\alpha}_{Np/l}D\right)$$

etc.

If we take the ratio of the first two reflections we have

$$\frac{A_1}{A_2} = 2\exp\left(2\bar{\alpha}_{Np/l}D\right) \qquad (1)$$

If we let g_{max} be the maximum of the g function then the amplitudes of these signal can all be written as shown. The ratio of the first two reflection amplitudes is then given by Eq. (1)

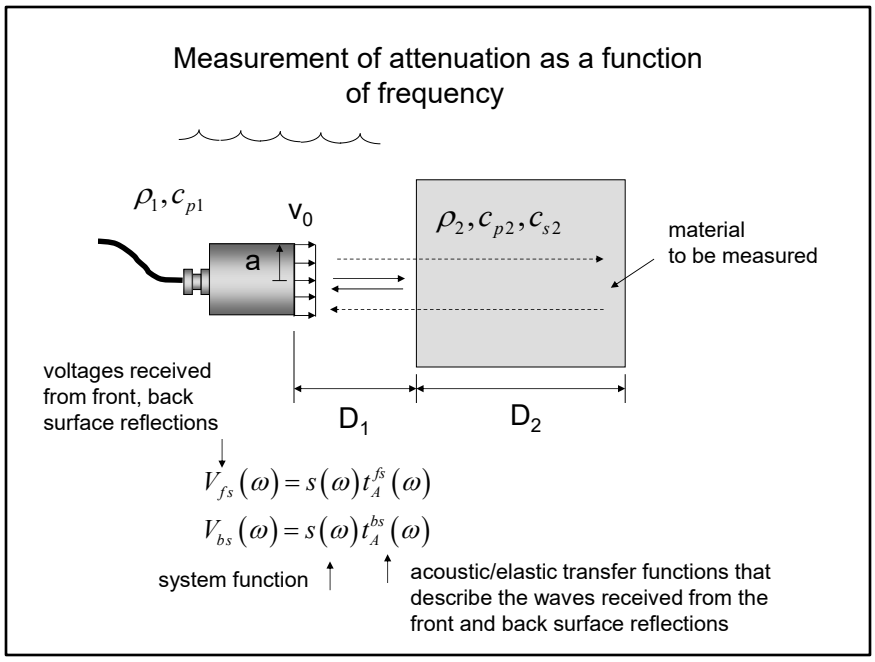
Now

$$20\log_{10}\left(\frac{A_1}{A_2}\right) = 20\log_{10}(2) + 20\log_{10}\left[\exp\left(2\bar{\alpha}_{Np/l}D\right)\right]$$
$$= 6\,dB + (2D)\,20\log_{10}(\mathrm{e})\,\bar{\alpha}_{Np/l}$$
$$= 6\,dB + (2D)\,\bar{\alpha}_{dB/l}$$

which gives

$$\bar{\alpha}_{dB/l} = \frac{20\log_{10}\left(\frac{A_1}{A_2}\right) - 6\,dB}{2D}$$

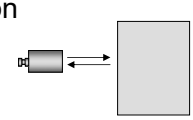
Assuming we measure this ratio we then can directly obtain an expression for the average attenuation. Note that because of the spherical wave spreading there is always a 6 dB decrease in the signal amplitude so that any material attenuation is seen in any changes greater than that 6 dB amount.



Now, consider how we might find a more detailed measurement of the **attenuation as a function of frequency**. We will again use a pulse-echo setup where we place both a transducer and a block of material (whose attenuation we want to measure) in a fluid and measure the front and back signals from the block [Note: here we are only considering waves that have been reflected once from the front or back surfaces.]. We have seen previously that in the frequency domain, these reflected signals can be written as the product of a system function and an acoustic/elastic transfer function that describes the waves present in this setup.

front surface transfer function

$$t_A^{fs}(\omega) = \tilde{t}_A^{fs}(\omega)\exp\left[-2\alpha_w(\omega)D_1\right]$$



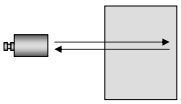
ideal transfer function model

$$\tilde{t}_A^{fs}(\omega) = \tilde{D}_p\left(k_{p1}a^2/2D_1\right)R_{12}\exp\left(2ik_{p1}D_1\right)$$

plane wave reflection coefficient

back surface transfer function

$$t_A^{bs}(\omega) = \tilde{t}_A^{bs}(\omega)\exp\left[-2\alpha_w(\omega)D_1 - 2\alpha_{p2}(\omega)D_2\right]$$



ideal transfer function model

$$\tilde{t}_A^{bs}(\omega) = \tilde{D}_p\left(k_{p1}a^2/2\bar{D}\right)T_{12}R_{21}T_{21}\exp\left(2ik_{p1}D_1 + 2ik_{p2}D_2\right)$$

plane wave reflection, transmission coefficients (normal incidence)

For ideal (non-attenuating) media we can actually model these acoustic/elastic transfer functions (see the next side) and then multiply them by the corresponding attenuation terms for both the water (whose attenuation is known) and the solid. Note that these models contain plane wave reflection and transmission coefficients at normal incidence.

$$\tilde{D}_p(u) = 2\left[1 - \exp(iu)\left\{J_0(u) - i\,J_1(u)\right\}\right]$$

diffraction coefficient

In the back surface signal, the distance $\bar{D}$ appearing in the diffraction correction is:

$$\bar{D} = D_1 + \frac{c_{p2}}{c_{p1}} D_2$$

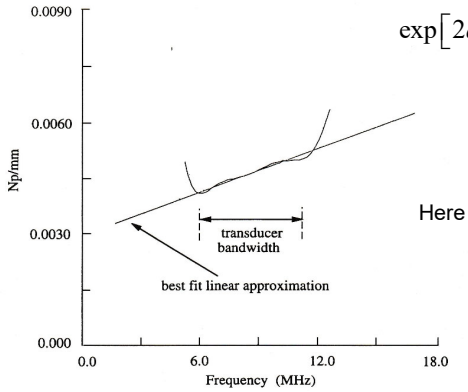
The acoustic/elastic transfer functions also contain diffraction coefficient functions that are in terms of Bessel functions J_0 and J_1.

$$\left|\frac{V_{bs}(\omega)}{V_{fs}(\omega)}\right| = \left|\frac{\tilde{t}_A^{bs}(\omega)}{\tilde{t}_A^{fs}(\omega)}\right| \exp\left[2\alpha_{p2}(\omega) D_2\right]$$

measure model

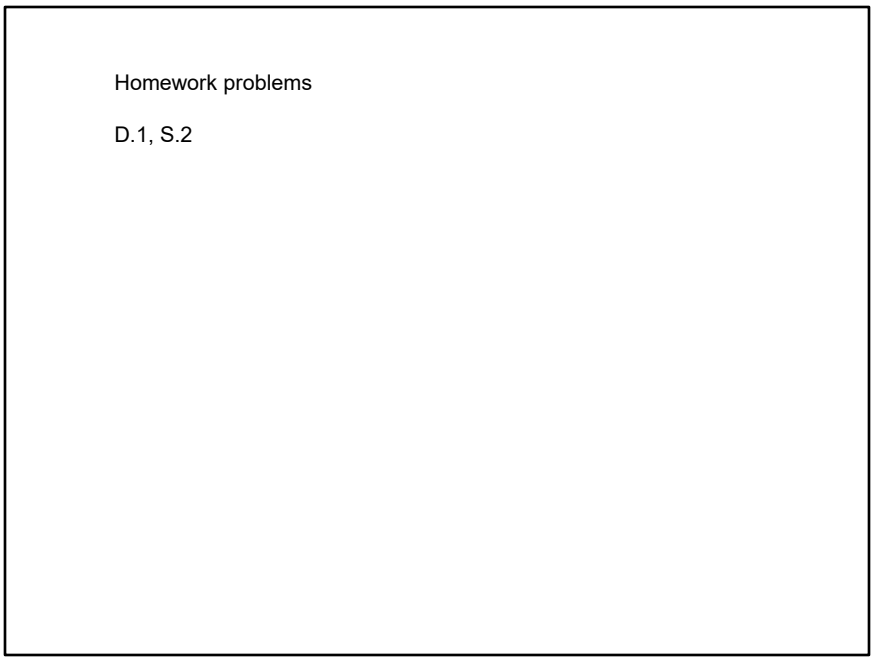
usually fit the attenuation to a simple polynomial function of frequency over the bandwidth of the system

$$\exp\left[2\alpha_s(\omega) D_2\right] = \left|\frac{V_{bs}(\omega)}{V_{fs}(\omega)}\right| \left|\frac{\tilde{t}_A^{fs}(\omega)}{\tilde{t}_A^{bs}(\omega)}\right|$$



Here is an example of such a fitting

If we measure the ratio of the magnitude of these front and back surface response, we know from our models the ratio of the acoustic/elastic transfer functions, so we can write an expression for the frequency dependent attenuation we seek. Normally, we obtain this attenuation by fitting it to a simple polynomial function of frequency.



Homework problem from Appendix D and a special problem.

Remarks:

To measure attenuation, it is not necessary to measure the system function (it cancels out).

Setup shown is for measurement of P-wave attenuation. S-wave attenuation measurements must be done in a different setup.

This approach is ad-hoc. Actual mechanisms of attenuation are rather complex.

For high attenuation, wave speed as well as amplitude is affected (material dispersion).

Here are some remarks on the method, which are self-explanatory.

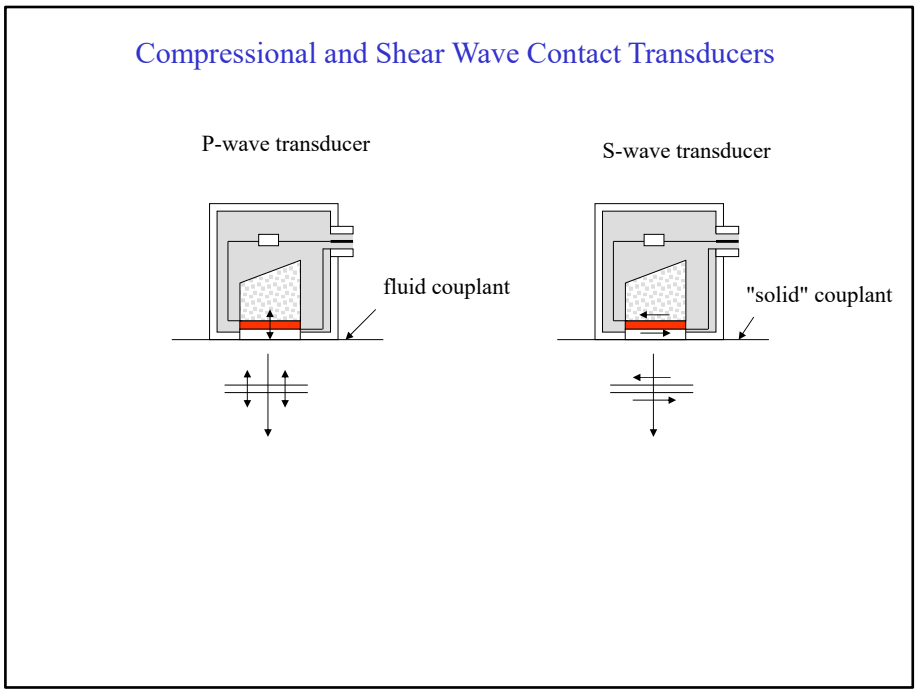
Waves Used in NDE

The waves we have been describing so far are bulk waves. There are a number of other types of waves used in NDE testing which we will also describe here.

Learning Objectives

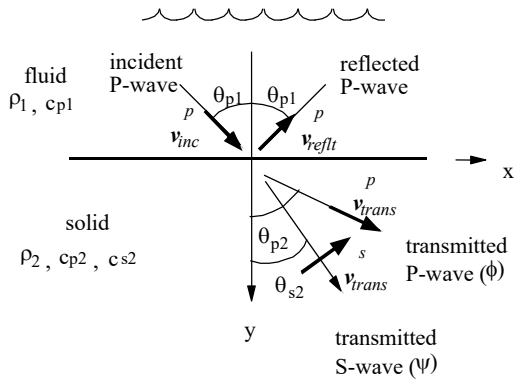
Bulk Compressional, Shear Waves
Rayleigh (surface) waves
Lamb (plate) waves

We will examine some ways that bulk waves are used in NDE and consider briefly some other types of waves such as surface waves and plate waves.



There are two types of contact transducers. In a P-wave transducer the transducer crystal experiences an expansion and contraction type of motion and transmits a corresponding P-wave with similar motions into the underlying elastic solid. To maintain good coupling to the material a thin fluid couplant layer is normally used. In contrast, an S-wave transducer crystal experiences shearing motions which are transmitted as an S-wave. In this case we must have a semi-permanent type of coupling such as glue between the transducer and the part to transmit this shearing motion. Thus, unlike a P-wave transducer, we cannot move the transducer around on the surface.

Mode Conversion

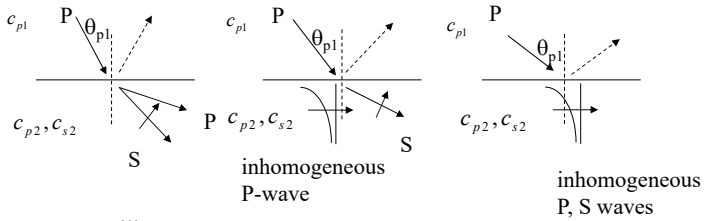


Generalized Snell's law:

$$\frac{\sin\theta_{p1}}{c_{p1}} = \frac{\sin\theta_{p2}}{c_{p2}} = \frac{\sin\theta_{s2}}{c_{s2}}$$

When waves are incident at an oblique angle to an interface one generates both compressional and shear waves whose angles are both controlled by Snell's law. Here we see an example of such plane waves in an immersion setup

Critical Angles



(1) (2) (3)

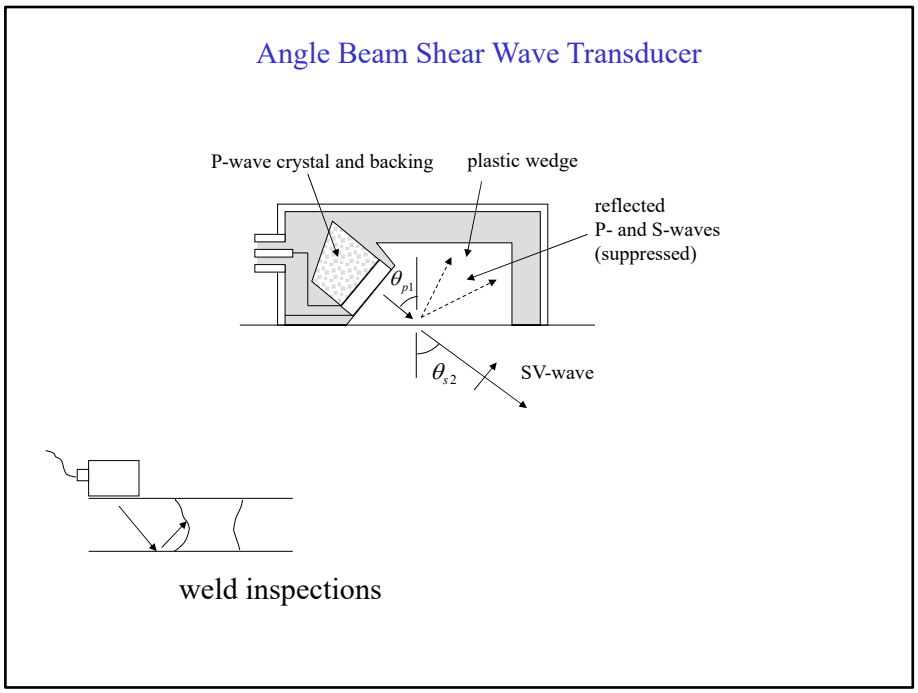
$$\theta_{p1} < \sin^{-1}\left(\frac{c_{p1}}{c_{p2}}\right) \qquad \sin^{-1}\left(\frac{c_{p1}}{c_{p2}}\right) < \theta_{p1} < \sin^{-1}\left(\frac{c_{p1}}{c_{s2}}\right) \qquad \theta_{p1} > \sin^{-1}\left(\frac{c_{p1}}{c_{s2}}\right)$$

$$\left(\theta_{CR}\right)_1 = \sin^{-1}\left(\frac{c_{p1}}{c_{p2}}\right) \qquad \left(\theta_{CR}\right)_2 = \sin^{-1}\left(\frac{c_{p1}}{c_{s2}}\right)$$

first critical angle

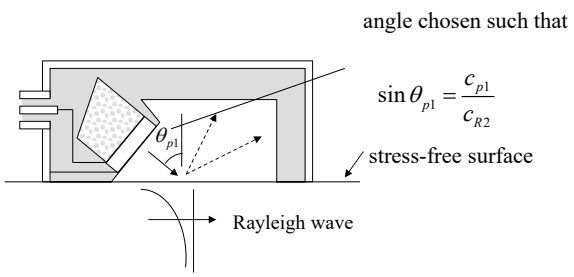
second critical angle

In oblique incidence problems involving elastic solids there are two critical angles present. For a fluid-solid interface case shown we first see the case (1) of the transmitted waves when we are below the first critical angle. In this case both P- and S-waves are transmitted into the solid. However, above the first critical angle the transmitted P-wave becomes an inhomogeneous wave (see (2)) that propagates along the interface and decays exponentially away from that interface, leaving only a transmitted S-wave to propagate in the solid. Above a second critical angle (case (3)) both the P- and S-waves become homogeneous waves traveling along the interface, and no transmitted wave exists in the solid. To have the first critical angle exist we must have $c_{p1} < c_{p2}$ and for the second critical angle to exist we must have $c_{p1} < c_{s2}$



We can use critical angle phenomena to generate an **angle beam shear wave transducer** where we place a P-wave crystal at an angle on a low speed wedge made of a material such as Lucite. This crystal generates primarily a P-wave in wedge which is transmitted as only an SV-wave into the underlying part if the angle is above the first critical angle but below the second critical angle. Again, a thin fluid couplant is used between the transducer and the part. Such angle beam shear wave transducers are commonly used for weld inspections where by moving the transducer back and forth on the surface we can scan different portions of the weld.

Rayleigh (Surface)Wave Transducer

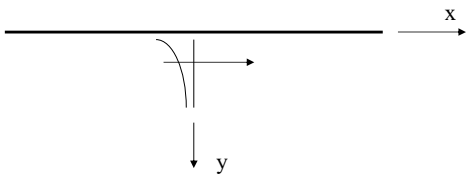


If we use an angle beam transducer setup but choose the incident angle so that it is slightly above the second critical angle, then we can generate a combination of inhomogeneous P- and S-waves that produce a Rayleigh wave which can travel for long distances along a stress-free surface. The Rayleigh wave speed (which we discuss shortly) is slightly smaller than the shear wave speed.

In the late 1800's Lord Rayleigh looked for a wave confined near the stress-free surface of an elastic solid of the form:

$$\phi = A\exp[-\alpha y]\exp[ik(x-ct)]$$

$$\psi = B\exp[-\beta y]\exp[ik(x-ct)]$$



To see how Rayleigh waves arise, consider a combination of inhomogeneous P- and SV-waves traveling along a stress-free surface. Rayleigh took potentials of the form shown above, having a wave speed, c, which at first is unspecified.

By satisfying the equations of motion

$$\nabla^2\phi - \frac{1}{c_p^2}\frac{\partial^2\phi}{\partial t^2} = 0$$

$$\nabla^2\psi - \frac{1}{c_s^2}\frac{\partial^2\psi}{\partial t^2} = 0$$

and the stress-free boundary conditions $\tau_{yy} = \tau_{xy} = 0$ on y = 0

Rayleigh found that the wave speed, c, must satisfy

$$\left(2 - c^2/c_s^2\right)^2 - 4\sqrt{1 - c^2/c_p^2}\sqrt{1 - c^2/c_s^2} = 0 \qquad (1)$$

c_P, c_S compressional and shear wave speeds

By satisfying both the equations of motion and the stress-free boundary conditions, Rayleigh found that wave speed, c, must satisfy the Rayleigh wave speed equation (1).

Rayleigh's equation

$$\left(2 - c^2 / c_s^2\right)^2 - 4\sqrt{1 - c^2 / c_p^2}\sqrt{1 - c^2 / c_s^2} = 0$$

There is always one real root of this equation, $c = c_R$

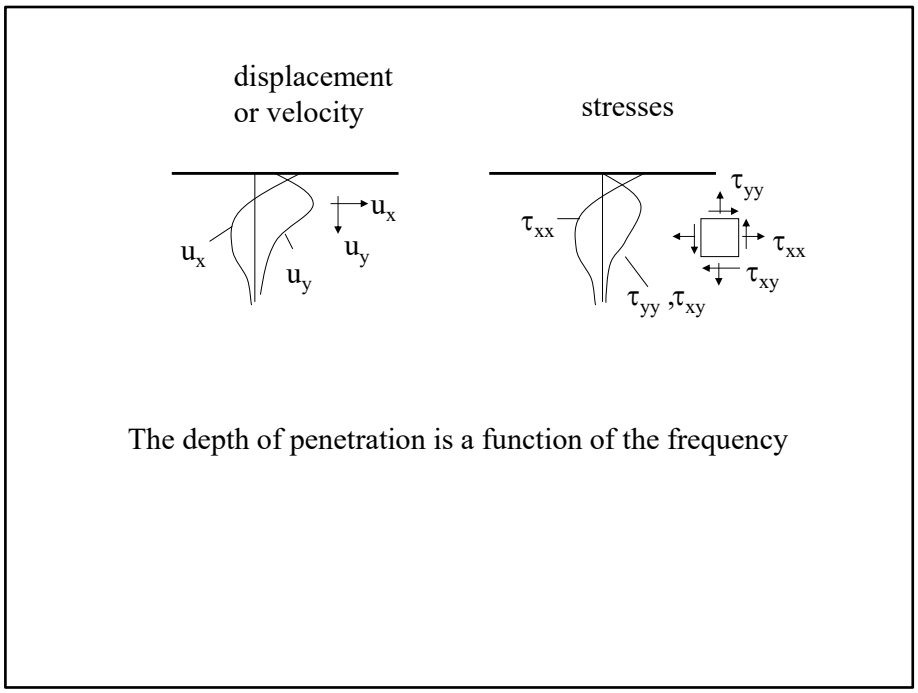
where $c_R < c_s$ A good approximation of this root is:

$$c_R \cong \frac{0.862 + 1.14\nu}{1 + \nu} c_s \qquad \nu \ldots \text{ Poisson's ratio}$$

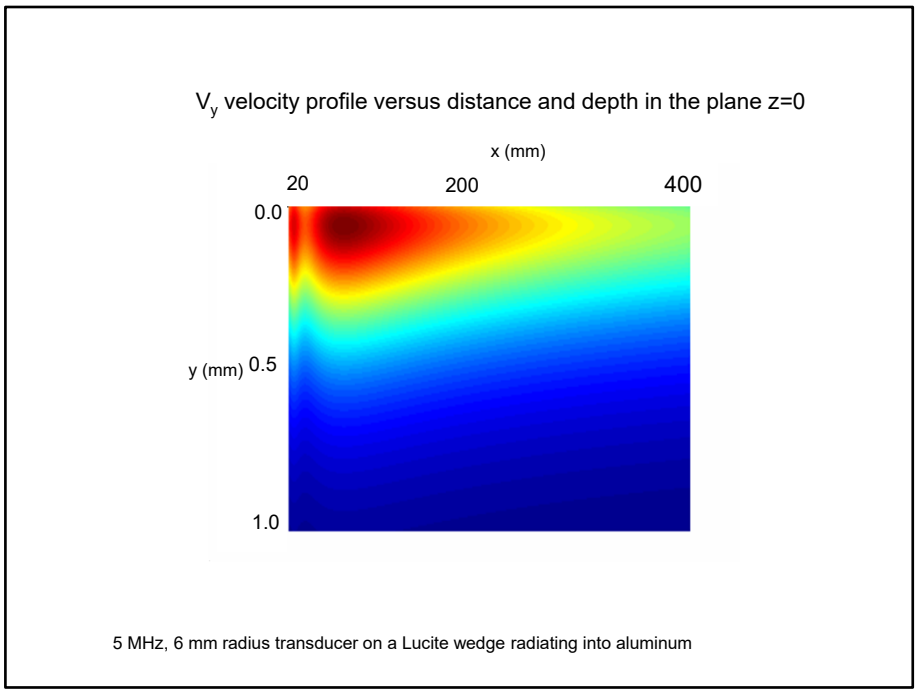
The Rayleigh wave travels about 90% of the shear wave speed

Rayleigh's equation always has a solution c =c_R which is slightly smaller than the shear wave speed. The simple approximate expression shown above can often be used to obtain the Rayleigh wave speed.

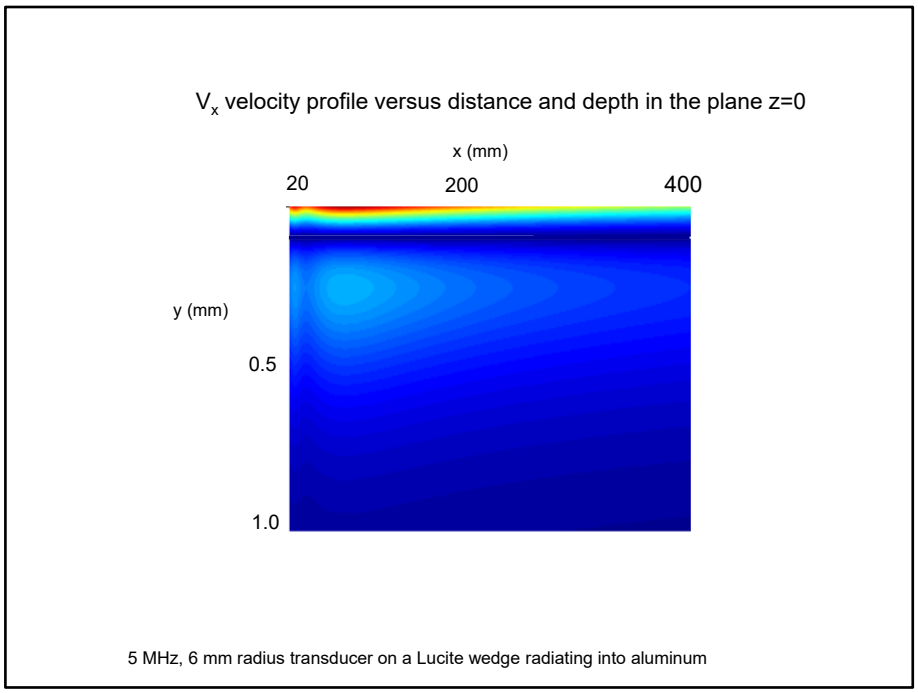
Rayleigh waves are very useful for applications such as inspecting the surface of parts for surface-breaking cracks since they are very sensitive to such flaws. Rayleigh waves can travel for long distances because, unlike bulk waves, they only spread out on the surface rather than spreading out through the entire volume of the part. They are, however, sensitive to surface conditions such as roughness.



Here are some plots of the displacements and stresses in a Rayleigh wave. The exact depth of penertration of the Rayleigh wave is a function of the frequency.

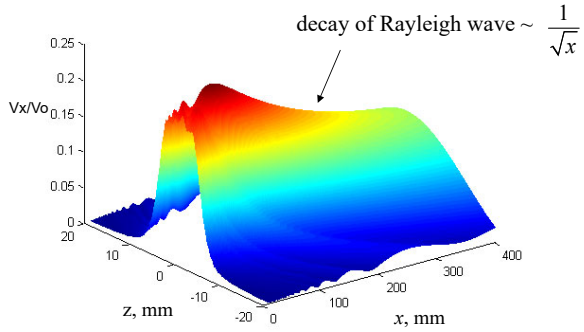


This, and the next three slides, will show some simulated wavefields generated by a Rayleigh wave transducer. Here, we see the vertical velocity profile in a vertical plane where red is high amplitude and blue is low amplitude



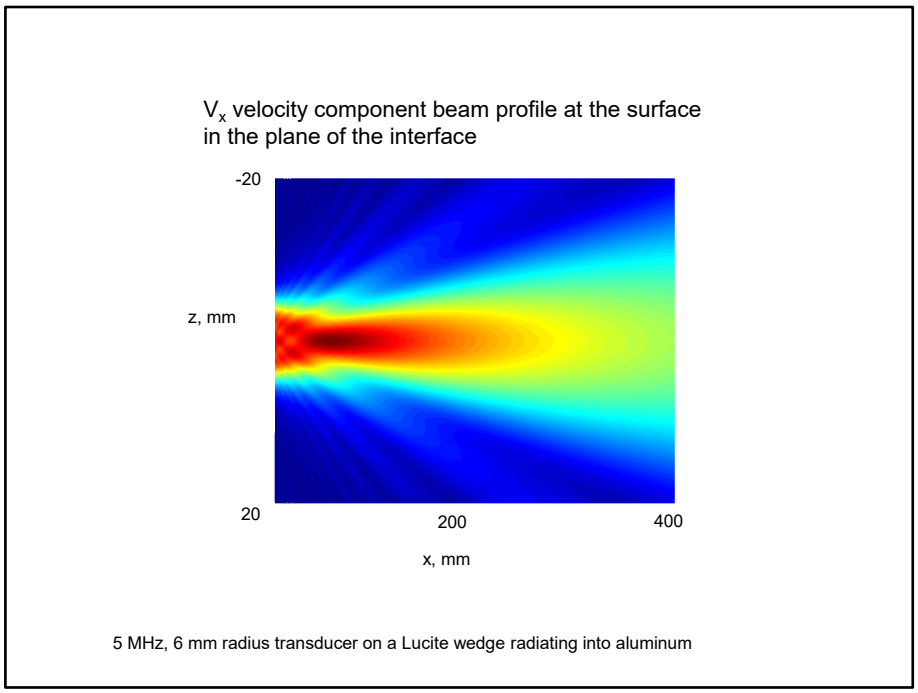
Here is the corresponding amplitude in the x-direction

2-D profile of the V_x velocity component at the surface



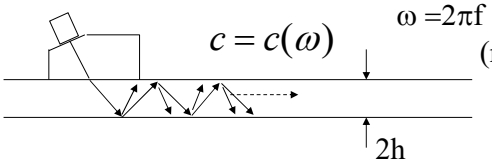
5 MHz, 6 mm radius transducer on a Lucite wedge radiating into aluminum

This is a plot of the x-velocity on the stress-free surface. At some distance from the transducer we see that the velocity profile is decreasing in amplitude like one over the square root of the distance, which is smaller than a spherical wave spreading which would be like 1/x.



Here we are looking directly down on the free surface at the x-velocity wavefield, showing the beam generated from the Rayleigh wave transducer.

Lamb (Plate) Waves

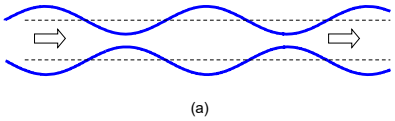


If one looks for solutions of the form

$$\phi = f(y)\exp[ik(x-ct)]$$
$$\psi = g(y)\exp[ik(x-ct)]$$

If we place an angle beam transducer on a thin plate, we will generate a complex set of multiply reflected P- and SV-waves. These can merge to form a wave, called a Lamb or plate wave, that travels in the thin plate with a frequency dependent wave speed. We can thus try to find solutions for such a wave in the plate which travels with a wave speed, c, and has some variations across the thickness of the plate.

then solutions of the following two types are found:

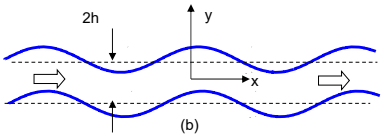


(a)

extensional waves

$$f = A\cosh(\alpha y)$$

$$g = B\sinh(\beta y)$$



(b)

flexural waves

$$f = A'\sinh(\alpha y)$$

$$g = B'\cosh(\beta y)$$

If we satisfy both the wave equations and the stress-free conditions on the surface of the plate we will find that we can have two types of plate wave disturbances called extensional waves and flexural (bending) wave where the functions f and g have the forms shown.

satisfying the boundary conditions $\tau_{yy} = \tau_{xy} = 0$
on $y = \pm h$ gives the Rayleigh-Lamb equations:

$$\frac{\tanh(\beta h)}{\tanh(\alpha h)} = \left[\frac{4\omega^2 \alpha\beta}{c^2\left(\omega^2 / c^2 + \beta^2\right)^2} \right]^{\pm 1}$$

\+ … extensional waves
\- … flexural waves

$$\alpha = \left|\frac{\omega}{c}\right| \sqrt{1 - \frac{c^2}{c_p^2}} \quad , \beta = \left|\frac{\omega}{c}\right| \sqrt{1 - \frac{c^2}{c_s^2}}$$

There are multiple solutions of these equations. For each solution the wave speed, c, is a different function of frequency. Each of these different solutions is called a "mode" of the plate.

The boundary conditions in particular give us the Rayleigh-Lamb equations for the wave speed, c, of the extensional and flexural waves. These equations have multiple solutions where each solution has a different dependency on the frequency. Each of these solutions is called a mode of the plate.

consider the extensional waves

$$\frac{\tanh\left[\,2\pi fh\sqrt{1/c^2-1/c_s^2}\,\right]}{\tanh\left[\,2\pi fh\sqrt{1/c^2-1/c_p^2}\,\right]}=\frac{4\sqrt{1-c^2/c_s^2}\sqrt{1-c^2/c_p^2}}{\left(2-c^2/c_s^2\right)^2}$$

If we let $kh=\dfrac{2\pi fh}{c}>>1$ (high frequency)

then both tanh functions are $\cong 1$

and we find $\left(2-c^2/c_s^2\right)^2=4\sqrt{1-c^2/c_s^2}\sqrt{1-c^2/c_p^2}$

so we just have Rayleigh waves on both stress-free surfaces:

If we examine the Rayleigh-Lamb equation for the extensional waves at high frequencies we find the same equation as for Rayleigh waves. In this limit the extensional waves are just Rayleigh waves propagating on both surfaces.

In contrast for $kh \ll 1$ (low frequency)

we find $\tanh(\alpha h) \cong \alpha h$

$\tanh(\beta h) \cong \beta h$

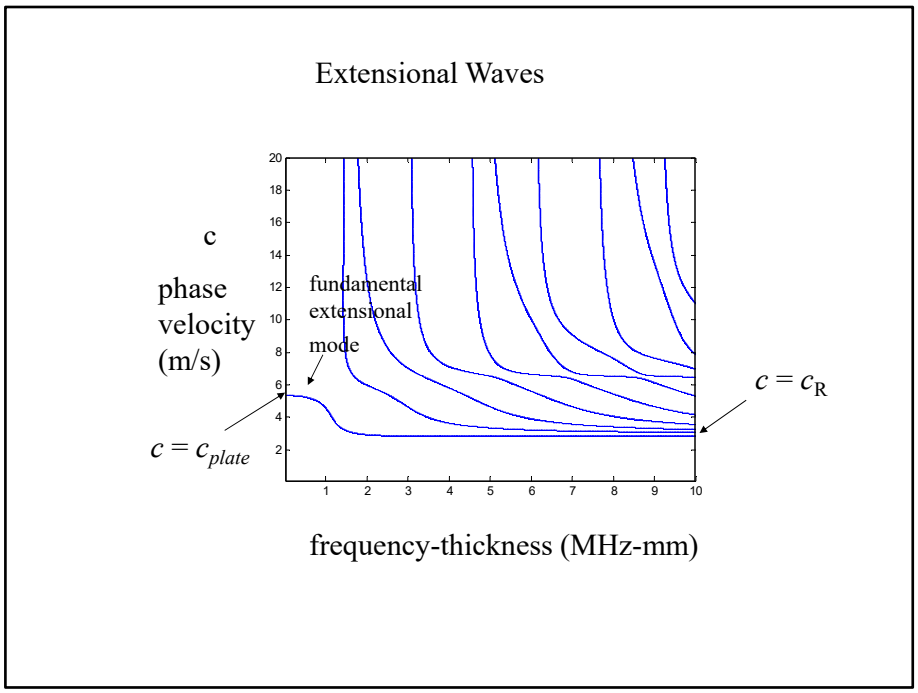
and the Rayleigh-Lamb equation reduces to

$$\left(2 - c^2 / c_s^2\right)^2 = 4\left(1 - c^2 / c_p^2\right)$$

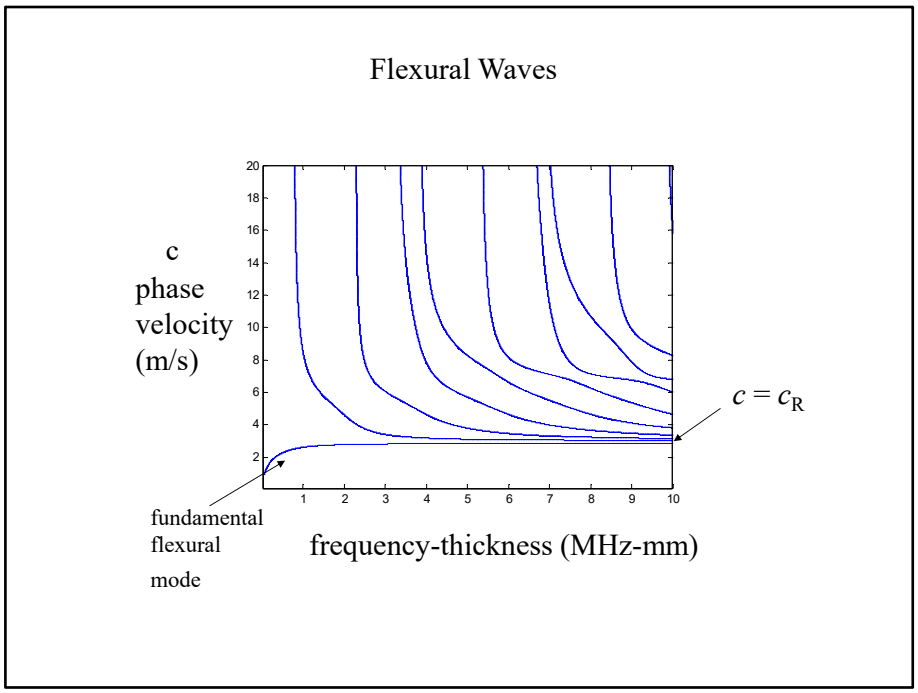
which can be solved for c to give

$$c = c_{plate} = \sqrt{\frac{E}{\rho\left(1 - \nu^2\right)}}$$

At low frequencies, the wave speed of extensional waves become a constant equal to the wave speed for P-waves in a plate.



Here we show the behavior of the first ten extensional modes in the plate as a function of frequency. Obviously there can be considerable complexity in these results but we will not discuss the details further here as our focus will be primarily on bulk waves.

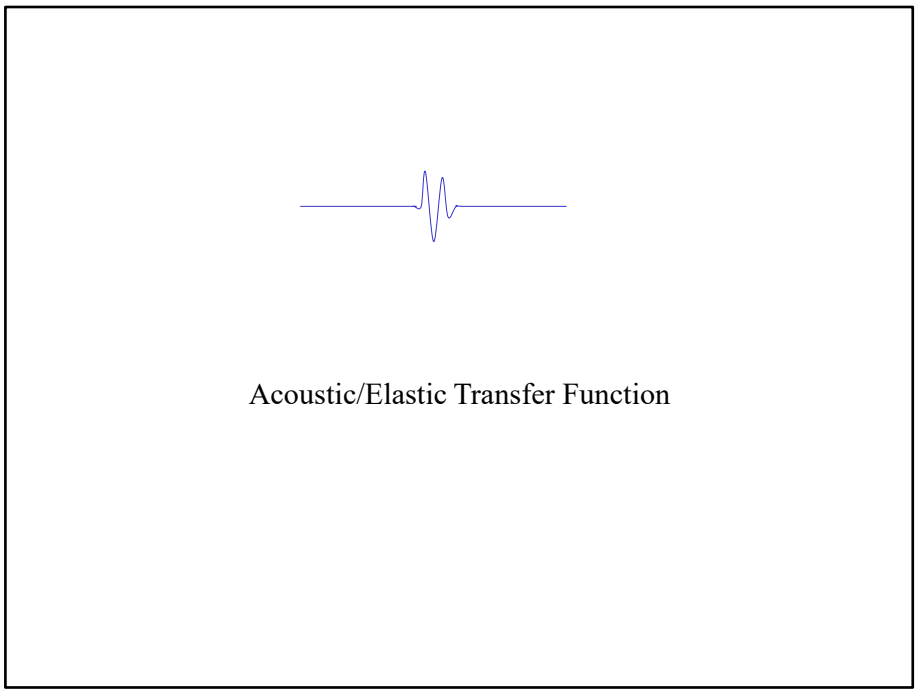


Here are the corresponding flexural plate mode curves. At low frequencies we see a much different behavior from that of extensional waves.

Homework problems

S.3, S.4, S.5

Special homework problems.



The acoustic/elastic transfer function describes the waves propagating in the fluid and solid media present in an NDE test. This is not a function we can measure so we need to be able to model the function.

Learning Objectives

Modeling the acoustic/elastic wave processes as a transfer function

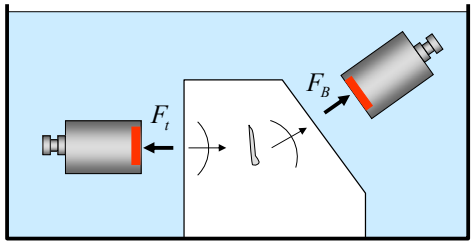
Blocked force concept

Special cases where the transfer function can be obtained analytically

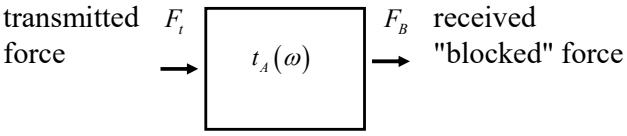
Effects of material attenuation

Here, we will define the acoustic/elastic transfer function and discuss the concept of the blocked force. We will examine some cases where one can actually determine the acoustic/elastic transfer function analytically and show such an example where attenuation is also included.

Acoustic/Elastic Transfer Function



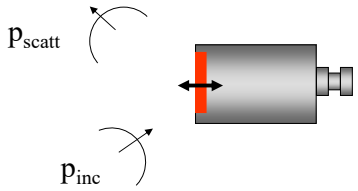
All the complex wave propagation and scattering processes occurring in an ultrasonic measurement system can be characterized in terms of an acoustic/elastic transfer function



The acoustic/elastic transfer function is defined as the ratio of forces at the receiving and sending transducers. The force at the sending transducer is just the driving compressive force on the face of that transducer but at the receiver the force used is the **blocked force**. We will examine those forces in a pitch-catch immersion setup but we can consider a contact testing pitch-catch setup in a similar manner.

The Blocked Force on Reception

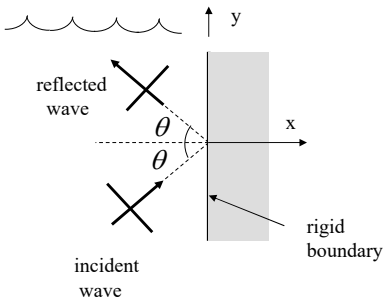
On reception the receiving transducer face moves and scattered waves are also generated



To see what the blocked force is consider a receiving transducer. Waves that are incident on that transducer will generate additional scattered waves.

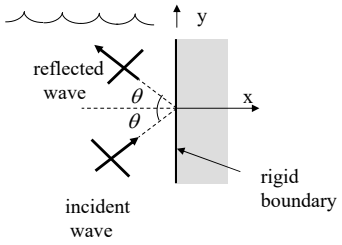
The Blocked Force on Reception

If the incident waves are treated locally like plane waves and diffraction effects are ignored, we can consider the interactions of the incident waves with the transducer face as those of plane wave with an infinite planar interface. Now, consider the case where the interface is held rigidly fixed:



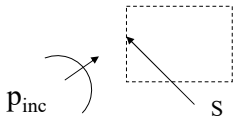
Although the incident and scattered waves are likely not planar, we will treat the interaction of the waves with the receiving transducer as plane waves. Consider the case when the face of the transducer is modeled as a rigidly fixed surface.

The Blocked Force on Reception



In this case it can be shown that at the interface the blocked force is just double the force due to the incident wave only as if the transducer were absent, i.e

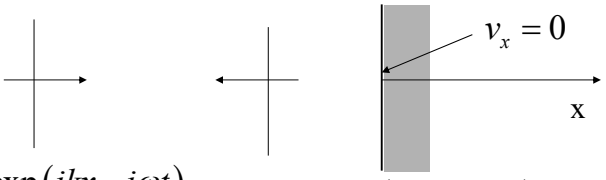
$$F_B(\omega) = 2F_{inc}(\omega) \quad \text{where} \quad F_{inc} = \iint_S p_{inc}\, dS$$



We will show that in this case the blocked force is just double that of the incident wave so we can compute the blocked force in this case only from the incident waves as if the transducer was absent.

The Blocked Force on Reception

proof for normally incident waves



$p_{inc} = P_i \exp(ikx - i\omega t)$ $p_{reflt} = P_r \exp(-ikx - i\omega t)$

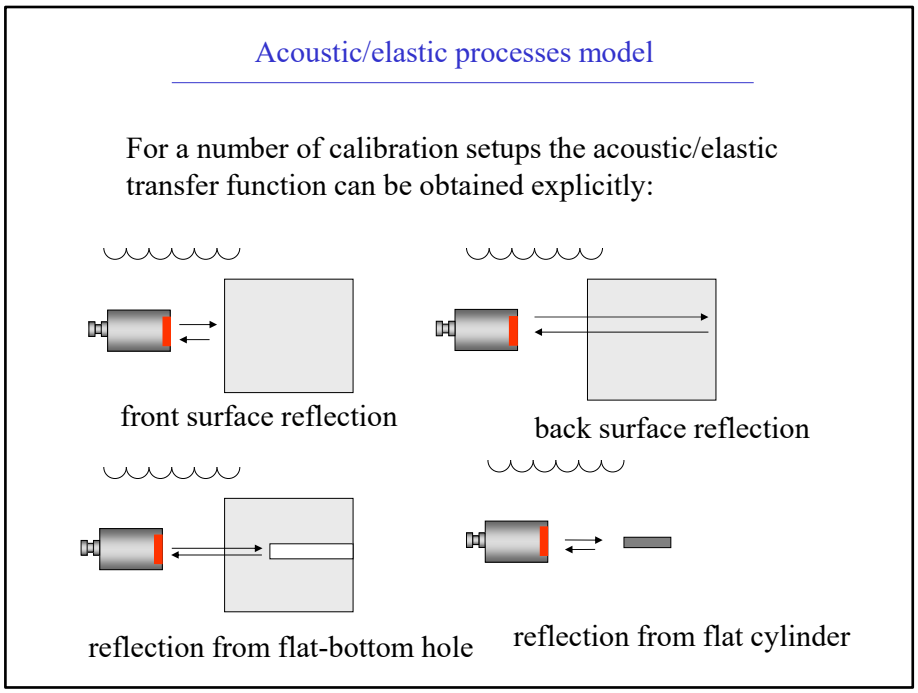
Equation of motion: $-\frac{\partial p}{\partial x} = -i\omega\rho v_x$ $p = p_{inc} + p_{reflt}$

from $v_x = 0$ at x = 0 $ikP_i - ikP_r = 0$

or $P_r = P_i$ (reflected pressure = inc. pressure)

Thus, also at x = 0 $F = F_{inc} + F_{reflt} = P_i S + P_r S = 2F_{inc}$

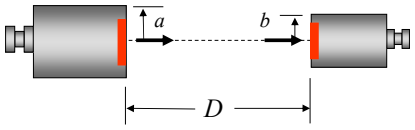
We have looked at plane wave interactions with an interface but now let us again examine the case where a plane wave is normally incident on a rigid (motionless) surface. Placing harmonic waves into Newton's law and using the boundary condition of no normal displacement, one finds the pressure amplitude of the reflected wave is just equal to that of the incident wave so that the total pressure is just double the incident pressure. The total force on the area of the receiving transducer face, therefore, will be just twice the incident force and this force is called the blocked force. Thus, in an ultrasonic setup if we can model the incident waves present at the receiving transducer coming from the sending transducer we can just double that force and use that result to calculate the acoustic/elastic transfer function.



In a number of setups we can model explicitly the acoustic/elastic transfer function. The ones shown here are all for pulse-echo setups.

Acoustic Transfer Function

Example: the acoustic/elastic transfer function of two aligned circular piston transducers of different radii can be found.



$$t_A(\omega) = \frac{2}{\pi a^2}\left\{\Theta \exp(ik_1 D) - 16a^2b^2 \cdot \int_0^{\pi/2} \frac{\sin^2 u \cos^2 u}{(a-b)^2 + 4ab\cos^2 u} \exp\left[ik_1 \sqrt{D^2 + (a-b)^2 + 4ab\cos^2 u} \right] du \right\}$$

$$\Theta = \begin{cases} \pi b^2 & a \geq b \\ \pi a^2 & b \geq a \end{cases}$$

We can also obtain the acoustic/elastic transfer function in this pitch/catch setup where two circular transducers are aligned along their central axes in a water bath. The explicit function is shown here in terms of an integral.

Acoustic Transfer Function

Special Cases

a = b :

$$t_A(\omega) = 2\left\{\exp(ik_1 D) - \frac{4}{\pi} \cdot \int_0^{\pi/2} \sin^2 u \exp\left[ik_1\sqrt{D^2 + 4a^2\cos^2 u}\right] du\right\}$$

At the high frequencies found for most NDE transducers the above integral can be evaluated approximately to yield:

$$t_A(\omega) = 2\exp(ikD)\left[1 - \exp\left(ika^2/D\right) \cdot \left\{J_0\left(ka^2/D\right) - iJ_1\left(ka^2/D\right)\right\}\right]$$

Here we show the special case where the two transducers are of the same size. In this case, at the high frequencies found in most ultrasonic NDE transducers the integral can be approximated in terms of Bessel functions to yield an explicit expression for the acoustic/elastic transfer function.

Acoustic Transfer Function

Other special cases:

a >> b : $$t_A(\omega) = 2\frac{b^2}{a^2}\left\{\exp[ikD] - \exp\left[ik\sqrt{D^2 + a^2}\right]\right\}$$

a << b : $$t_A(\omega) = 2\left\{\exp[ikD] - \exp\left[ik\sqrt{D^2 + a^2}\right]\right\}$$

D >> a, b : $$t_A(\omega) = -ika^2\frac{\exp(ik_1 D)}{D}$$

Here are also some special cases where we can obtain analytical expressions for the transfer function.

Acoustic Transfer Function

Attenuation of ultrasound in the fluid can be added to this model:

$$t_A(\omega) \rightarrow t_A(\omega)\exp\left[-\alpha(\omega)D\right]$$

$\alpha(\omega)$ … attenuation coefficient D … distance traveled

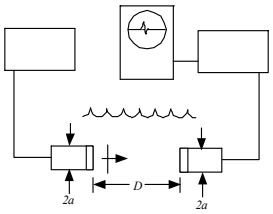
For water $\alpha = 25.3\times10^{-6} f^2$ 1/mm

f… frequency in MHz

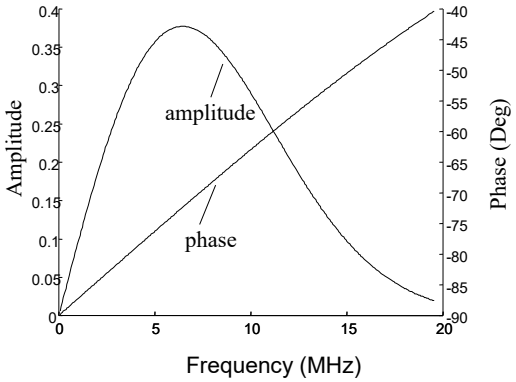
All the results we have shown for the acoustic/elastic transfer functions were for an ideal (i.e. non-attenuating) fluid media. However, as we showed previously we can easily include an attenuation factor to account for losses, and for water, in particular, we can describe the attenuation coefficient.

Acoustic Transfer Function

Example acoustic transfer function (including fluid attenuation effects)



D = 444 mm
a = 3.175 mm

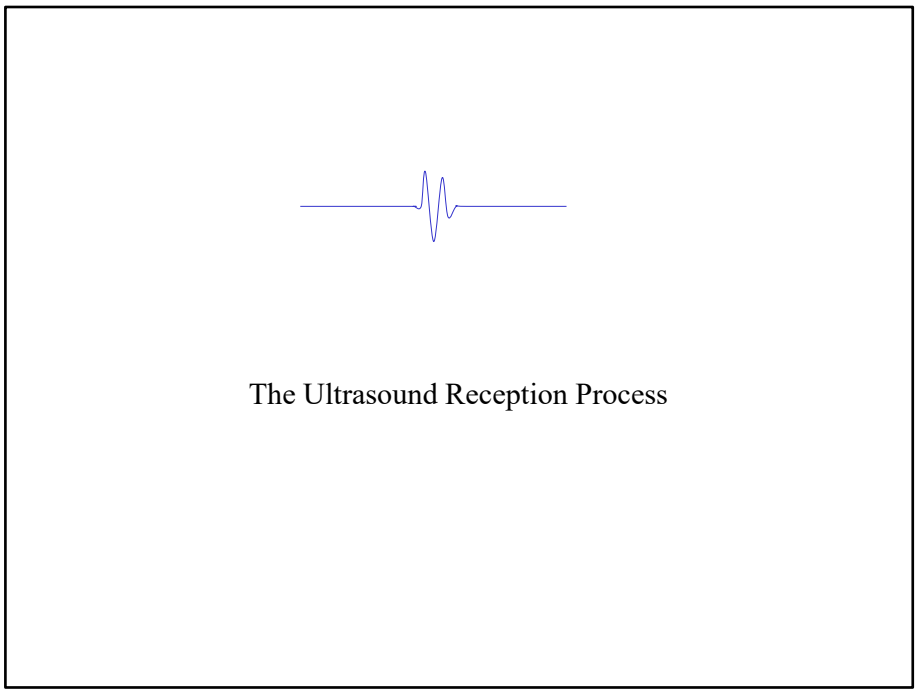


Here is an example calculation for the acoustic/elastic transfer functions for a pair of NDE immersion transducers where attenuation of the water is included in the calculations.

Acoustic Transfer Function

Goldstein, A., Gandhi, D.R., and W.D. O'Brien," Diffraction effects in hydrophone measurements," IEEE Trans. Ultrasonics, Ferroelectrics, and Freq. Control, 45, 972-979, 1998.

Here is a reference on some of the issues discussed here. There are many others as well.



Here, we will describe all the elements of the reception process where the acoustic/elastic waves are transformed into a measured output voltage.

Learning Objectives

Blocked force as a "lumped" source on reception

Thevenin equivalent transducer/sources on reception

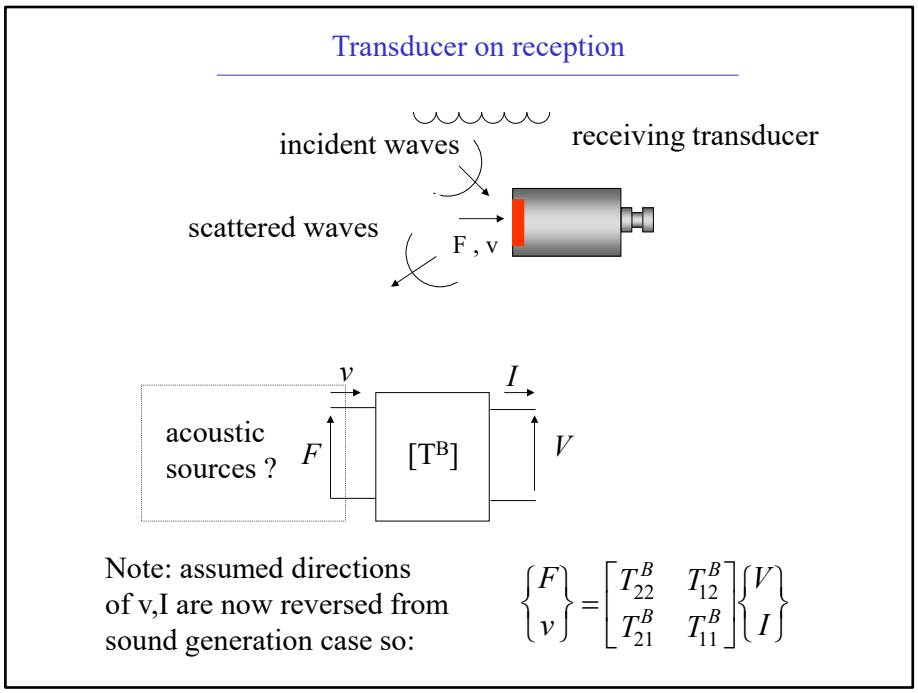
Measurement of receiving cabling properties

Characterization of receiver as impedance, gain and measurement of these receiver parameters

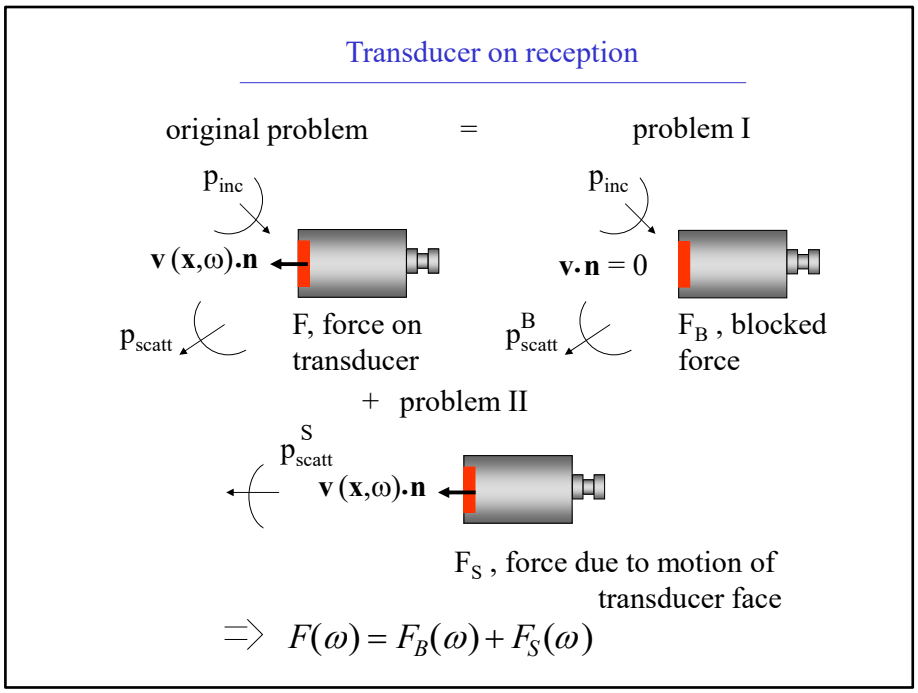
Reception transfer function

Ultrasonic system model

We will examine why the blocked force is important in the reception process and examine how this force acts as part of a Thevenin equivalent model of the receiving transducer and its acoustic sources. Again, we will examine the cabling and discuss how to model the receiver. All of these elements will then be combined into a complete model of the reception process.



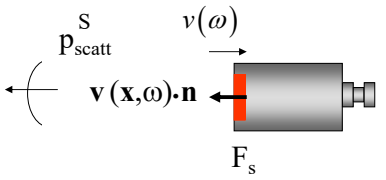
Like the sending transducer the receiving transducer can be modeled as a two port system but where now the acoustic port is the driving input port and the electrical is the output. Also note that the directions of the velocity and current are now changed from the sending case. These changes means that the elements of the transfer matrix used to characterize the transducer as a transmitter are in different locations in the transfer matrix when the transducer is used as a receiver, as shown. To make this model useful we must describe how the acoustic waves acting on the receiver can be defined in terms of a lumped source parameter.



Consider a problem where a set of waves are incident on a receiving transducer, producing some scattered waves as well as some motion of the face of the transducer. We can consider this problem as the superposition of problem I, where the face of the transducer is held fixed, and problem II where the incident waves are absent and the transducer has the motion of the face of the original problem. The force on the transducer face for problem I is the blocked force, F_B, while the force due to the motion of the face in problem II we will call F_s. Obviously the total force is the sum of these two forces.

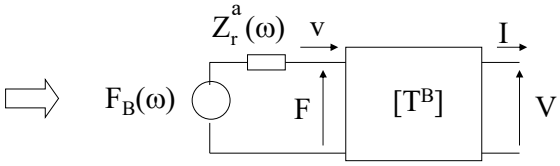
Transducer on reception

problem II : transducer is acting as a transmitter



$$F_S(\omega) = Z_r^a(\omega)\left(-v(\omega)\right)$$

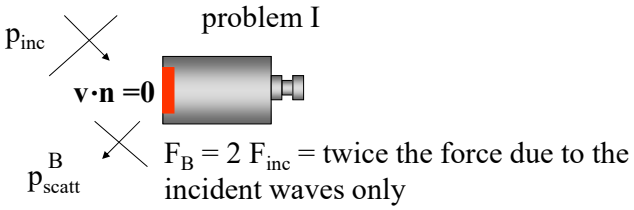
$$F(\omega) = F_B(\omega) - Z_r^a(\omega)v(\omega) \qquad (1)$$

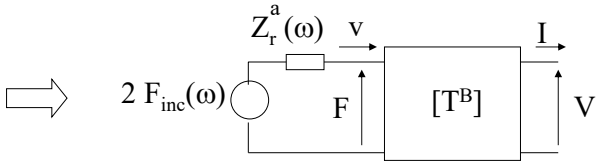


In problem two the waves generated are due only to the motion of the transducer face so this is identical to the case where the transducer is used as a transmitter except now the direction of the velocity is different. Thus, a minus sign appear when we relate the force and the velocity through the acoustic radiation impedance. The total force is then given by Eq. (1) and this relationship can be modeled as shown in the final figure where the blocked force acts as a source term in series with the acoustic radiation impedance at the input port. Thus, the blocked force is the natural force to calculate at the receiving transducer.

Transducer on reception

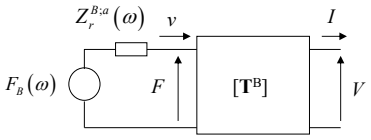
Note: many authors assume the incident and scattered waves can be treated as plane waves incident on a fixed planar surface. In this case, as we showed previously:



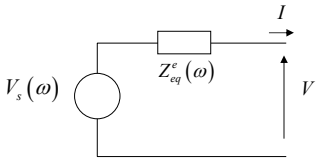


As we showed previously, if we model the incident and scattered waves at the receiver locally as plane waves, then the blocked force is just twice the force in the incident waves when the receiving transducer is absent so we can replace the blocked force by twice this incident force in our model.

Transducer on reception



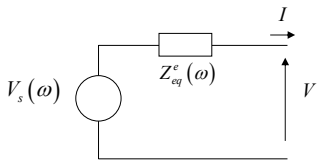
Thevenin equivalent
reduced transducer and acoustic sources model



We can replace the acoustic sources and transducer by a Thevenin equivalent system consisting of only a voltage source and an electrical impedance.

Transducer on reception

reduced transducer and acoustic sources model



under open circuit conditions

$$\begin{Bmatrix} F \\ v \end{Bmatrix} = \begin{bmatrix} T^B_{22} & T^B_{12} \\ T^B_{21} & T^B_{11} \end{bmatrix} \begin{Bmatrix} V^\infty \\ 0 \end{Bmatrix} \quad \Rightarrow \quad \begin{aligned} F(\omega) &= T^B_{22}(\omega) V^\infty(\omega) \\ v(\omega) &= T^B_{21}(\omega) V^\infty(\omega) \end{aligned}$$

but, recall $F(\omega) = F_B(\omega) - Z^{B;a}_r(\omega) v(\omega)$

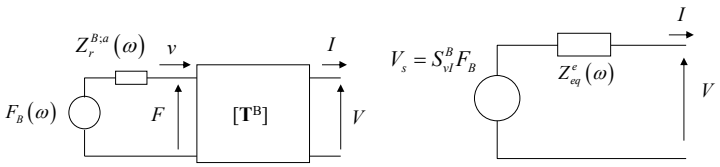
open-circuit voltage

so $$\frac{V^\infty(\omega)}{F_B(\omega)} = \frac{1}{T^B_{22}(\omega) + Z^{B;a}_r(\omega) T^B_{21}(\omega)} = S^B_{vI}(\omega) \quad \Rightarrow \quad V_s = V^\infty = S^B_{vI}(\omega) F_B(\omega)$$

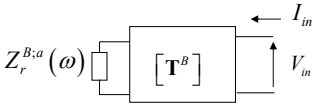
We can show that the open circuit voltage is just the blocked force times the sensitivity factor we discussed previously when the transducer is used as a transmitter.

Transducer on reception

reduced transducer and acoustic sources model



To get the impedance, remove the blocked force source and measure $Z_{eq}^e = \frac{V_{in}}{I_{in}}$:

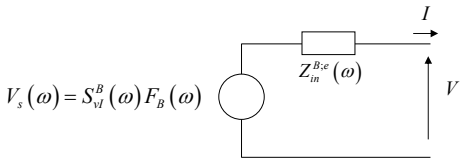


but this is the same configuration as during sound generation so $Z_{eq}^e = Z_{in}^{B;e}$

We can get the Thevenin equivalent impedance by removing the blocked force source, but this just yields the same configuration as when the transducer is being used as a transmitter so the electrical impedance here is the same as the one calculated when the transducer is being used as a transmitter.

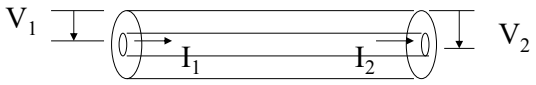
Transducer on reception

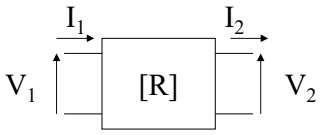
reduced transducer and acoustic sources model



Thus, the final Thevenin equivalent circuit is as given, showing that the transducer impedance and sensitivity characterize the transducer completely when used as both a transmitter or receiver.

Receiving cable – transfer matrix model



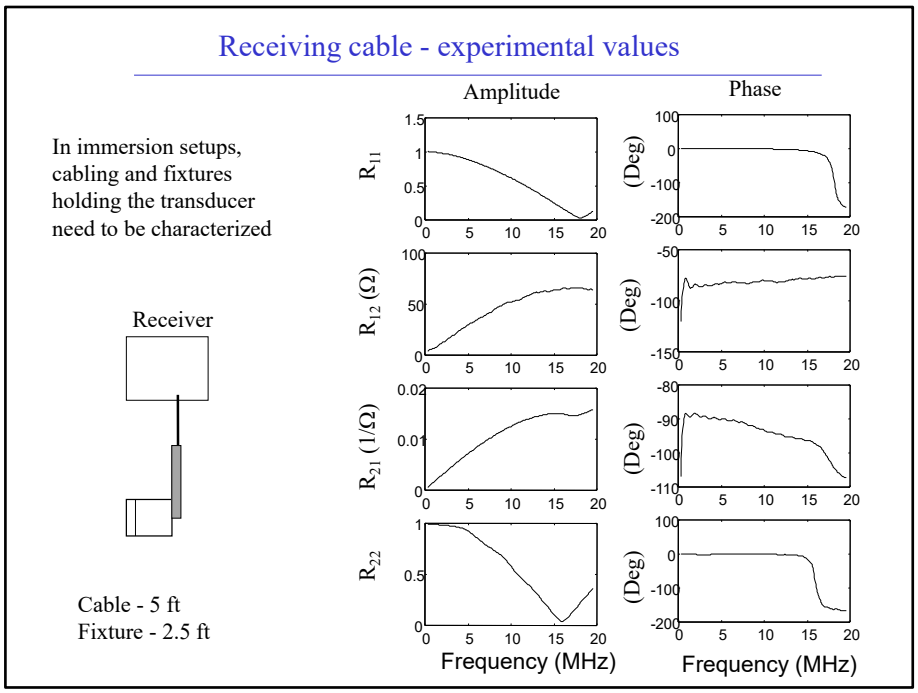


transmission line model

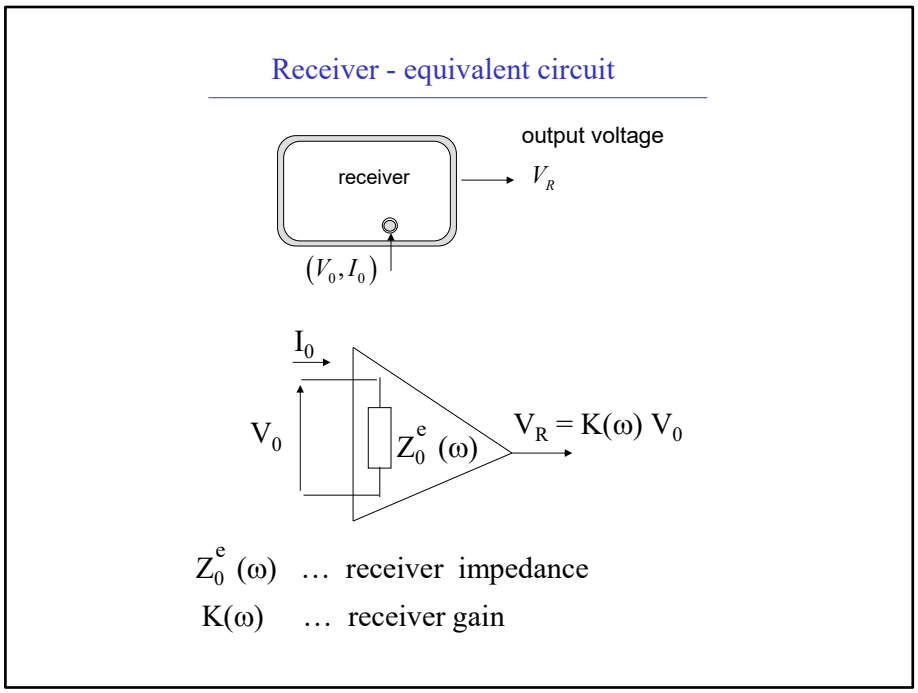
$$\begin{Bmatrix} V_1 \\ I_1 \end{Bmatrix} = \begin{bmatrix} \cos(k_c l_c) & -iZ_0^e \sin(k_c l_c) \\ -i\sin(k_c l_c)/Z_0^e & \cos(k_c l_c) \end{bmatrix} \begin{Bmatrix} V_2 \\ I_2 \end{Bmatrix}$$

$R_{11} = R_{22}$ so changing directions on currents does not change [R]

For the receiving cable we can again use the same two port system transfer matrix whose elements can be obtained with electrical measurements as shown previously.

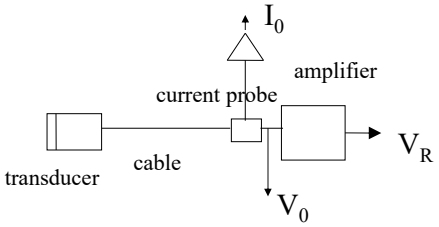


In immersion measurements there may be internal cables in fixtures holding the transducer as well as coaxial cables to the receiver but we can measure the transfer matrix of such combinations directly in the same manner, as shown here. The behavior of the results are in general agreement with what we expect from the transmission line model.



The receiver part of the pulser will have some characteristic input electrical impedance and will amplify the signal. The amplification we can characterize by a frequency dependent gain factor. We will not include any low or high pass filtering in our receiver model since in quantitative NDE measurements we normally do not enable those filters. Such filters, however, can easily be added to our model if desired.

Receiver - experimental



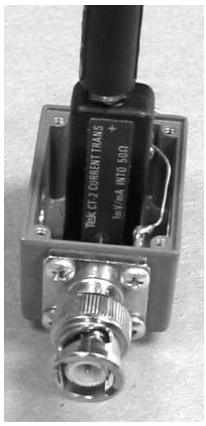
$$Z_0^e(\omega) = V_0(\omega) / I_0(\omega)$$

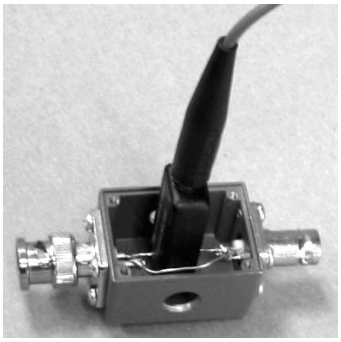
$$K(\omega) = V_R(\omega) / V_0(\omega)$$

If we measure the input current and voltage as a function of frequency at the receiver input and the receiver voltage output we can calculate both the impedance and the gain.

Ultrasonic System Components

current probe



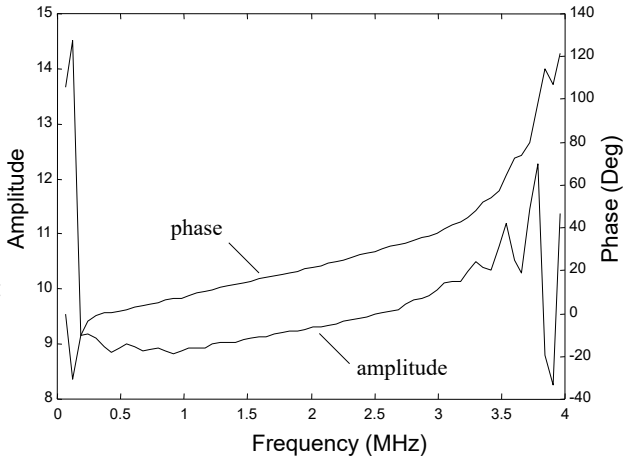


The voltage is measured by simply tapping the voltage at a T-adapter while the current is measured by stripping off a small section of cable and clamping a commercially available current probe onto the exposed inner conductor. This assembly is placed in a small box as shown.

Receiver - experimental

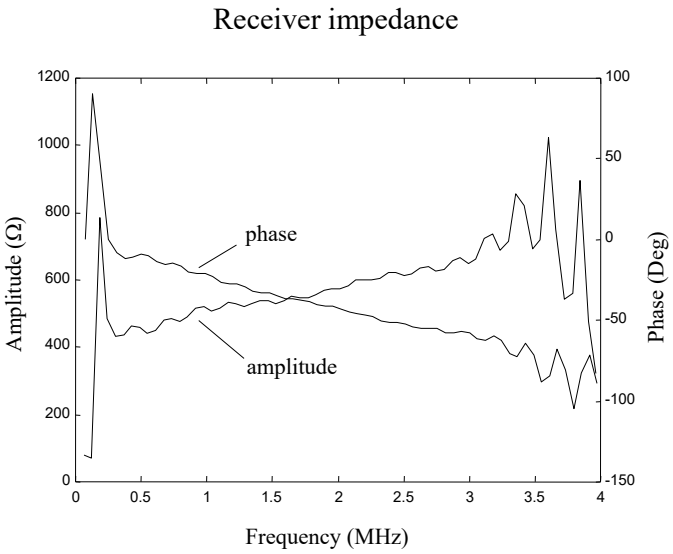
Receiver gain measured with a 2.25 MHz transducer

Receiver settings:
gain: 20 dB
attenuation: 0 dB



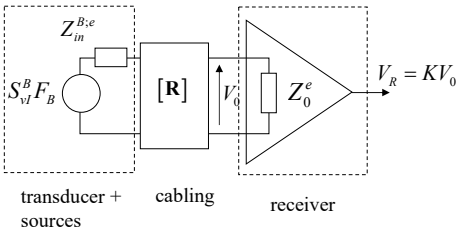
Shown are some measurements of the amplitude and phase of the receiver gain function at a particular gain and attenuation setting. The electrical inputs were generated in an ultrasonic setup where a 2.25 MHz transducer was connected to the receiver. Thus, the measurements here are only good for the bandwidth of this transducer. If we need to have a wider response we could use a different wide-band source or a different transducer.

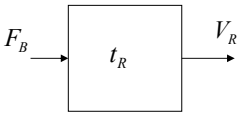
Receiver - experimental



Here is a corresponding impedance measurement of the receiver.

Reception process model





$$t_R(\omega) = \frac{V_R(\omega)}{F_B(\omega)} = \frac{K Z_o^e S_{vl}^B}{\left(Z_{in}^{B;e} R_{11} + R_{12}\right) + \left(Z_{in}^{B;e} R_{21} + R_{22}\right) Z_o^e}$$

We can combine all the elements of the reception process into a single transfer function for reception as shown.

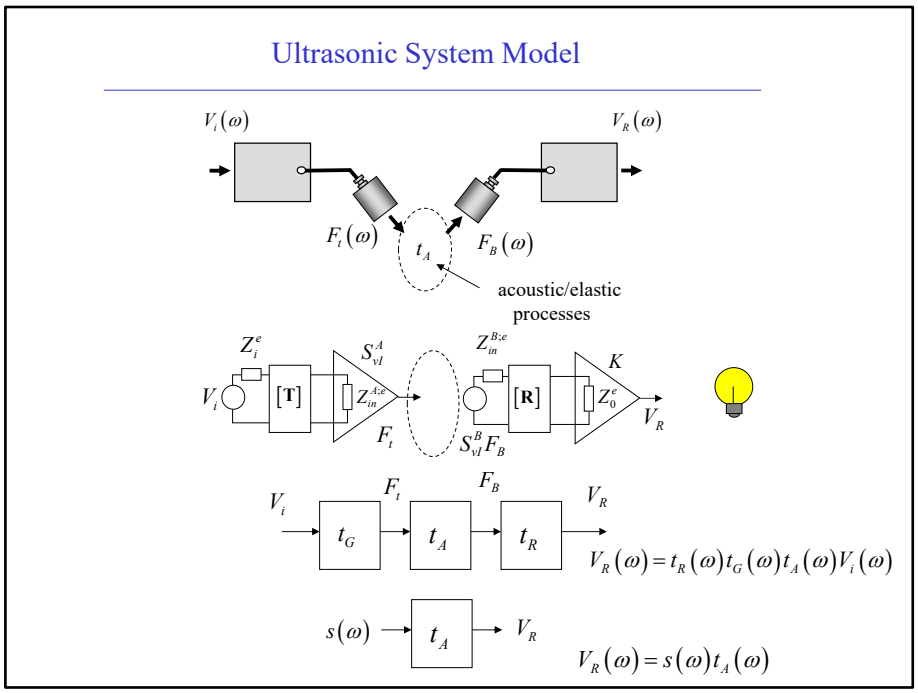
Generation and Reception process models

Both the generation and reception process transfer functions depend only on the transducer sensitivities and their electrical impedances:

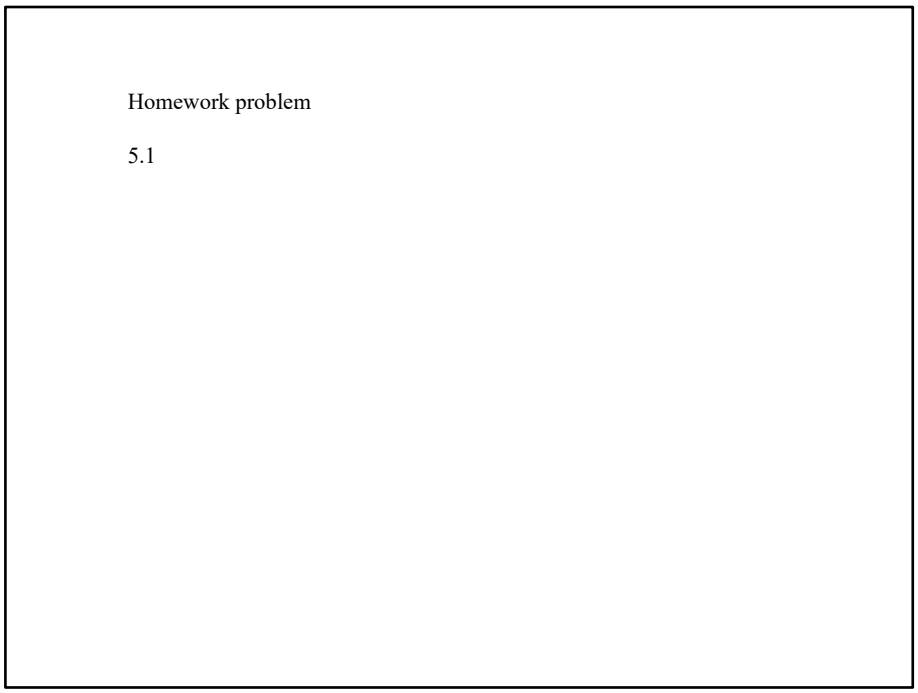
$$t_R(\omega) = \frac{V_R(\omega)}{F_B(\omega)} = \frac{K Z_o^e S_{vI}^B}{\left(Z_{in}^{B;e} R_{11} + R_{12}\right) + \left(Z_{in}^{B;e} R_{21} + R_{22}\right) Z_o^e}$$

$$t_G(\omega) = \frac{F_t(\omega)}{V_i(\omega)} = \frac{Z_r^{A;a} S_{vI}^A}{\left(Z_{in}^{A;e} T_{11} + T_{12}\right) + \left(Z_{in}^{A;e} T_{21} + T_{22}\right) Z_i^e}$$

Here we show the complete transfer functions for both the sending and receiving parts of an ultrasonic measurement system for a pitch-catch immersion setup. We see that the transducers impedance and sensitivity are the only parameters we need to completely characterize the role of the transducers in a measurement so we do not need to know their transfer matrix components explicitly.



We now have complete models of both the sound generation and sound reception parts of an ultrasonic measurement system, leading to the three LTI systems shown. If we combine the generation and reception process transfer functions with the Thevenin equivalent voltage source into a system function then we have an even simpler model consisting of a single LTI system where the system function acts as the input and the transfer function is the acoustic/elastic transfer function. Thus, if we have a practical way to measure the system function and if we can model the acoustic/elastic transfer function, we can predict the measured output voltage.



Homework problem from Chapter 5.

Transducer Characterization

Here we will examine the measurements needed to fully characterize the transducer(s) used in an ultrasonic measurement system.

Learning Objectives

Pulse-echo experimental determination of impedance, sensitivity of commercial transducers

Experimental determination of effective radii and focal lengths of commercial transducers

The transducer parameters we will need to obtain experimentally are the transducer electrical impedance, the transducer sensitivity, and the effective radius and focal length (if focused) of the transducer.

Transducer impedance, sensitivity

$$t_G(\omega) = \frac{F_t(\omega)}{V_i(\omega)} = \frac{Z_r^{A;a} S_{vI}^A}{\left(Z_{in}^{A;e} T_{11} + T_{12}\right) + \left(Z_{in}^{A;e} T_{21} + T_{22}\right) Z_i^e}$$

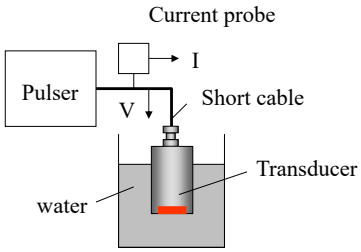
$$t_R(\omega) = \frac{V_R(\omega)}{F_B(\omega)} = \frac{K Z_o^e S_{vI}^B}{\left(Z_{in}^{B;e} R_{11} + R_{12}\right) + \left(Z_{in}^{B;e} R_{21} + R_{22}\right) Z_o^e}$$

To fully characterize these generation and reception transfer functions, we need to be able to obtain the transducers input electrical impedance and their sensitivities

Again, recall that the transducer electrical impedance and sensitivity are the two key parameters we need to find to determine the transducer's role in the measurement process.

Transducer impedance

Measurement of the transducer input impedance, Z_{in} :



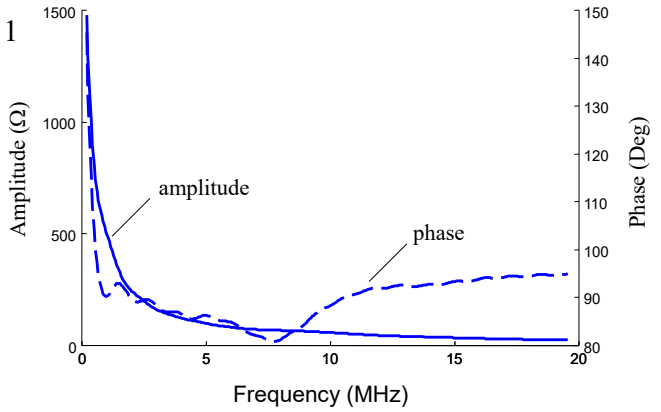
$$Z_{in}(\omega) = \frac{V(\omega)}{I(\omega)}$$

The electrical impedance is easy to obtain if we simply connect a transducer and short cable to the pulser and measure the voltage and current in the cable before any reflected signals are received. Taking the Fourier transform of those signals then we find the impedance directly.

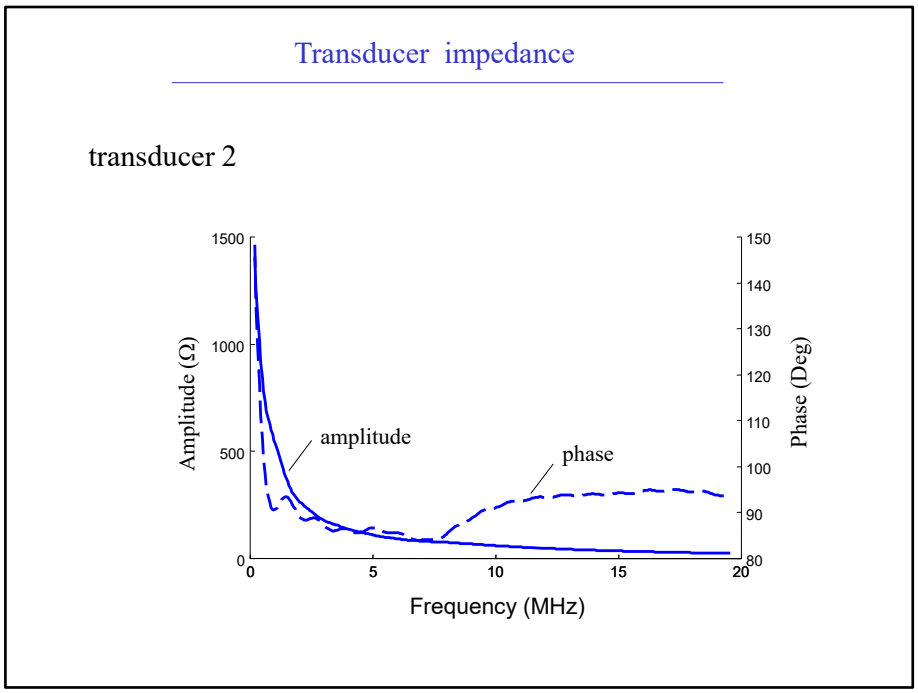
Transducer impedance

Example impedance measurements of two 5 MHz transducers:

transducer 1



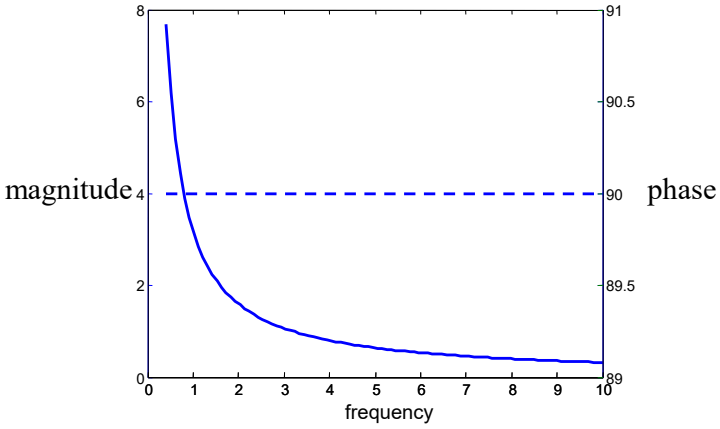
We will show impedances measurements for two 5 MHz transducers. Here are the results for the first transducer.



Here are the results for the second transducer.

Transducer impedance

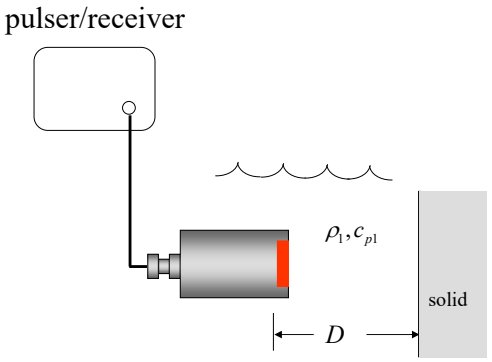
Impedance of a capacitor, $Z = 1/(-i\omega C)$



Both of those transducers have a frequency dependent behavior that looks much like that of a capacitor. Here is a capacitor's electrical impedance. This is not surprising since a transducer is a piezoelectric crystal that is plated on both faces.

Transducer sensitivity

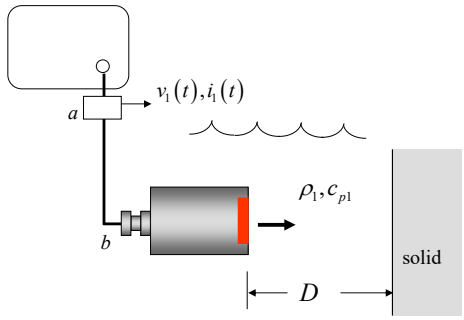
Pulse-echo measurement setup for determining sensitivity



To determine the sensitivity we can use a pulse-echo immersion setup where we place the transducer in a water bath at normal incidence to the flat face of a solid block.

Transducer sensitivity

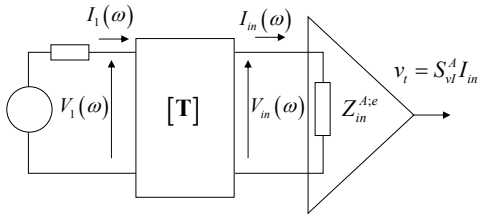
1. measure voltage and current when transducer is radiating into the fluid but before any reflected waves have arrived



First, we measure the voltage and current at the pulser/receiver before any reflected waves have arrived.

Transducer sensitivity

2. do FFT of the measured voltage and current and relate to the voltage and current at the transducer by compensating for the cabling



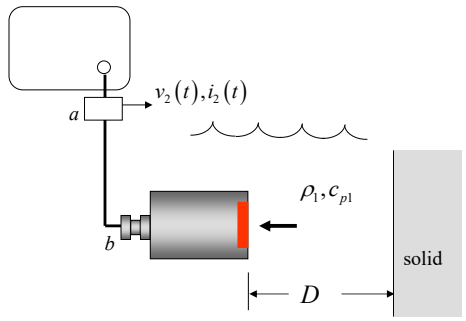
$$\begin{Bmatrix} V_{in} \\ I_{in} \end{Bmatrix} = \frac{1}{\det[\mathbf{T}]} \begin{bmatrix} T_{22} & -T_{12} \\ -T_{21} & T_{11} \end{bmatrix} \begin{Bmatrix} V_1 \\ I_1 \end{Bmatrix}$$

Note: then the impedance of the transducer is just $Z_{in}^{A;e} = \frac{V_{in}}{I_{in}}$

We do the FFT of those voltage and current measurements and then, for a known cable, compensate for the cabling effects to obtain the corresponding voltage and current at the input port of the transducer. The ratio of this voltage and current again is just the electrical impedance, but here we do not have to use a short cable to obtain that impedance.

Transducer sensitivity

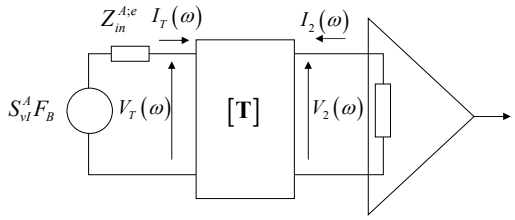
3. measure voltage and current generated by the waves reflected from the surface of the block:



Then we measure the received voltage and current at the pulser/receiver generated by the waves reflected from the block

Transducer sensitivity

4. do FFT of the measured voltage and current and relate to the voltage and current at the transducer electrical port by compensating for the cabling



$$\begin{Bmatrix} V_T \\ -I_T \end{Bmatrix} = \frac{1}{\det[\mathbf{T}]} \begin{bmatrix} T_{22} & -T_{12} \\ -T_{21} & T_{11} \end{bmatrix} \begin{Bmatrix} V_2 \\ I_2 \end{Bmatrix}$$

We then do an FFT of the measured reflected wave signals (called V_2 , I_2 here) and compensate for the cabling to get the voltage and current (V_T , I_T) at the transducer electrical port. Note the changes in sign consistent with the definitions of inputs and outputs and the assumed directions of the currents. Ideally, the diagonal elements of the cable would be identical and the cable would be completely reciprocal but we have allowed for small differences in the actual measurements.

Transducer sensitivity

5. From these measurements and a knowledge of the acoustic/elastic transfer function for this setup we can obtain the sensitivity of the transducer from:

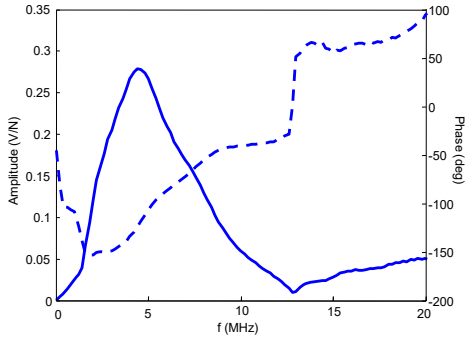
$$S_{vI}^{A} = \sqrt{\frac{V_{in} I_T + V_T I_{in}}{t_A Z_r^{A;a} I_{in}^2}}$$

[Note: in all these division processes in the frequency domain, a Wiener filter is used to desensitize the process to noise]

One can show that the transducer sensitivity can be obtained from these measurements since we can model the acoustic/elastic transfer function for this setup and the acoustic radiation impedance of the transducer. We will not show those models here.

Transducer sensitivity

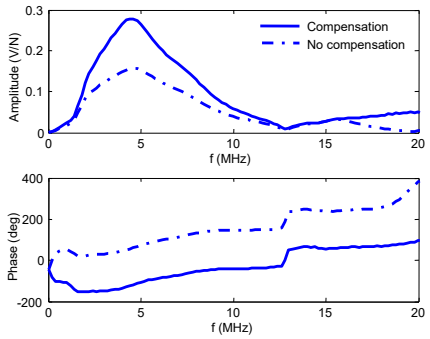
Example sensitivity calculated for a 5MHz planar transducer:



Here is an example sensitivity calculated for a 5 MHz transducer.

Transducer sensitivity

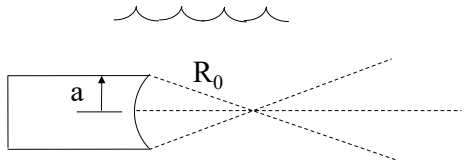
Cabling effects need to be accounted for in determining the sensitivity:



The compensations for cabling were important in these measurements as shown here, where the results when cabling effects are ignored are also shown. We could, of course, also eliminate those cabling effects by doing the measurements directly at the transducer's electrical port, but because this is an immersion setup those measurements would have to be taken under the water and it is more practical to do the measurements at the pulser/receiver.

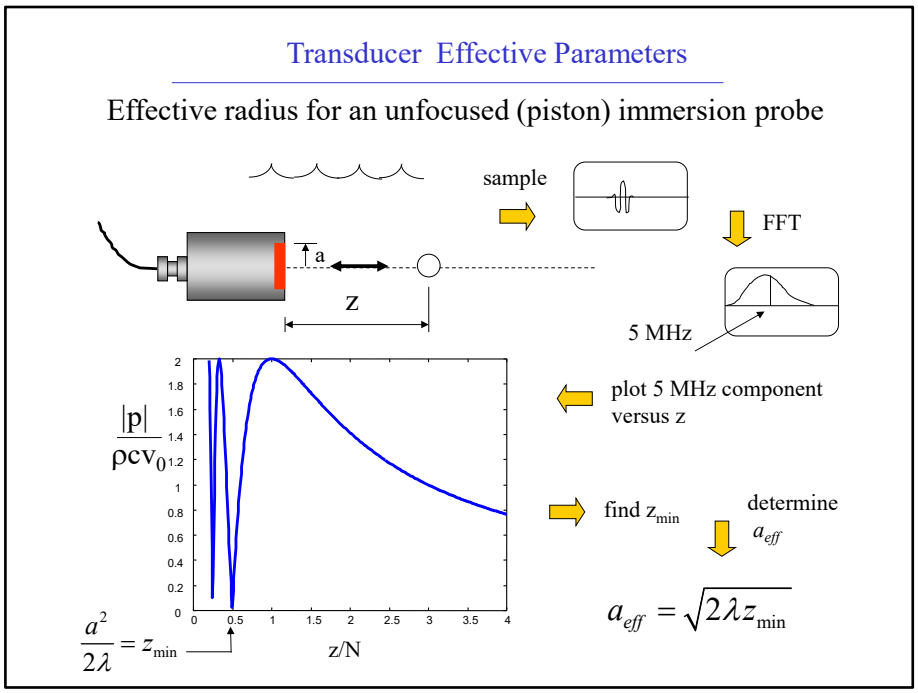
Transducer Effective Parameters

Use of manufacturer specifications for parameters such as transducer radius and focal length in transducer models do not always lead to good agreement with experimental measurements.



Thus, we need to determine experimentally "effective" values for these parameters.

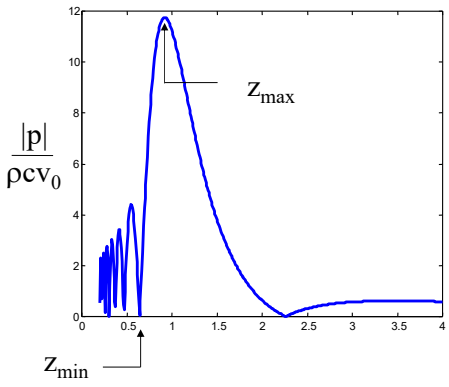
A manufacturer will specify quantities such as transducer radius and focal length for a commercial transducer but those values may not be able to be used directly in models of ultrasonic measurements without causing errors, which can sometimes be large. Thus, we need to be able to determine experimentally these parameters, which we will call **effective parameters**.



We can measure the effective radius of an unfocused circular transducer by placing a small reflector in a immersion tank and moving that reflector at different distances along the central axis of the transducer. A frequency domain model of the on-axis pressure distribution on the axis of the transducer shows that in the so-called near field of the transducer there are series of peaks and nulls in the pressure at a given frequency. Those pressure variations will cause the measured frequency components of the received voltage from the reflector also to vary. Thus, if one measures the time domain waveform received from the reflector at a given distance and does an FFT to get the frequency components of the signal, then one can look at the response at a given frequency (say 5MHz as shown above). If this same 5 MHz response is obtained when the reflector is at different distances, one can locate the distance, z_{min} , where the 5 MHz response has its last zero value. Since this distance is related to the radius of the transducer and the wavelength (and hence, the frequency) one can use it to predict an effective radius value. In principle this effective radius should not depend on the choice of frequency but experiments show that there can be some variation with frequency so if that is the case then one usually uses an average effective radius value calculated at different frequencies over the bandwidth of the transducer.

Transducer Effective Parameters

Effective radius and focal length for a spherically focused probe



$$a_{eff} = \sqrt{\frac{2\lambda z_{\min}\left(R_0\right)_{eff}}{\left(R_0\right)_{eff} - z_{\min}}}$$

$$\left(R_0\right)_{eff} = z_{\max}\left\{\frac{\pi - x}{\pi - x\left(z_{max}/z_{\min}\right)}\right\}$$

where x is the root of:

$$x\cos\left(x\right) = \frac{\pi - x\left(z_{\max}/z_{\min}\right)}{\pi - x}\sin\left(x\right)$$

For a spherically focused transducer we must determine both an effective radius and an effective focal length so we will need at least two measured values. Again, we can use the same setup just described and obtain the magnitude of the response at a given frequency for a reflector at different distances along the transducer central axis. A model of a spherically focused circular transducer shows that the transducer has an on-axis pressure response as shown above. Here, one measures the location of the large peak of the response (which is called the true focus) and the same last on-axis null. The model then also shows that we can obtain the effective radius and focal length values from those two measurements. The null location is usually easy to measure but the exact location peak is a bit more difficult to obtain so one can use a number of possible peak locations to try to determine those effective values that best fit the on-axis response at other points as well.

Transducer Effective Parameters

Effective transducer parameters determined experimentally:

Probes	Manufacturers Specs		Estimated Parameters		Center Frequency (MHz)
	focal length (cm)	radius (cm)	focal length (cm)	radius (cm)	
A	7.62	0.476	13.47	0.451	10
B	7.62	0.635	20.74	0.556	5
C	7.62	0.476	7.45	0.469	15

These values should be independent of frequency but in practice they do vary somewhat with the frequency component used in their determination.

Here are some example calculations of effective transducer parameters. These values were obtained by measuring the on-axis response at the center frequency of the transducer. Other frequencies may give somewhat different values. One can see that in some cases the results are quite different from the manufacturer's specifications.

Homework problem

6.1

Homework problem from Chapter 6.

The System Function

Here we will examine the system function and its experimental determination.

Learning Objectives

Combination of transfer functions and Thevenin equivalent pulser source term into a single factor – the system function

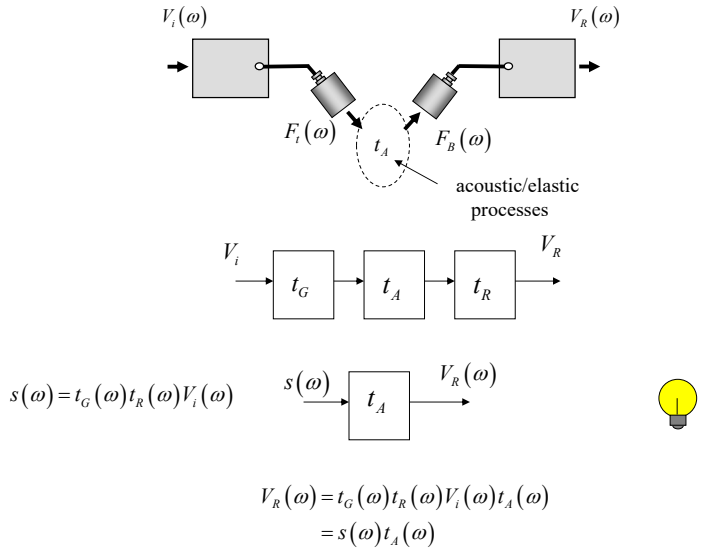
Relationship between the system function and the system "efficiency" factor

Experimental determination of the system function by deconvolution

Examples of synthesizing an entire ultrasonic measurement system and predicting the measured signals

One can collect the sound generation and sound reception transfer functions together with the Thevenin equivalent voltage source of the pulser into a single function called the **system function** which we can obtain experimentally through models and measurements . We will relate this system function to a closely related function called the **system efficiency factor** which has been used in earlier studies. We will give a number of examples of the use of a system function to predict the detailed characteristics of measured ultrasonic signals.

The System Function



$$s(\omega) = t_G(\omega)t_R(\omega)V_i(\omega)$$

$$V_R(\omega) = t_G(\omega)t_R(\omega)V_i(\omega)t_A(\omega)$$
$$= s(\omega)t_A(\omega)$$

Here we see a complete ultrasonic NDE system characterized by three transfer functions and a “reduced” model of the system where the sound generation and reception transfer functions are combined with the Thevenin equivalent voltage source to model the system as simply the acoustic/elastic transfer function where the system function serves as the driving input.

The System Function

$$s(\omega) = t_G(\omega)\, t_R(\omega)\, V_i(\omega) \quad \ldots \text{ system function}$$

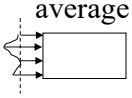
There is a relationship between $s(\omega)$ and the system "efficiency" factor, $\beta(\omega)$, [Schmerr, Fundamentals of Ultrasonic Nondestructive Evaluation, 2nd Ed.] :

$$V_R(\omega) = \beta(\omega)\, t_A^p(\omega) \qquad t_A^p(\omega) = \frac{p_{ave}(\omega)}{\rho c v_t(\omega)}$$

piston velocity

$v_t(\omega)$

average incident pressure



$p_{ave}(\omega)$

In earlier studies the acoustic/elastic transfer function was defined as the average incident pressure acting on the receiving transducer divided by the density and wave speed multiplied by the uniform velocity on the face of the transmitting transducer, as shown. When we relate the received voltage to this acoustic/elastic transfer function and a driving input, the driving input is called the **system efficiency factor**, which different from the **system function**. We can use either the system function or the system efficiency factor but we must be aware of the differences.

The System Function

$$V_R(\omega) = \beta(\omega) t_A^p(\omega) \qquad V_R(\omega) = s(\omega) t_A(\omega)$$

where $$t_A^p(\omega) = \frac{p_{ave}(\omega)}{\rho c v_t(\omega)}, \quad t_A(\omega) = \frac{F_B(\omega)}{F_t(\omega)}$$

For piston transducers, high frequency

$$F_B(\omega) = 2 p_{ave}(\omega) S_R$$
$$F_t(\omega) = \rho c v_t(\omega) S_T$$

S_R , S_T … areas of receiving and transmitting transducers

$$\Rightarrow \qquad s(\omega) = \frac{S_T}{2 S_R} \beta(\omega)$$

If we compare the system function and system efficiency factor and the corresponding acoustic/elastic transfer functions then for piston transducers and at high frequencies, where the acoustic radiation impedance is just a plane wave acoustic impedance, we see the system function and system efficiency are just proportional to each other.

The System Function

Two methods to find s(ω):

A. Measure all the components needed to obtain $t_R(\omega)$, $t_G(\omega)$, $V_i(\omega)$ by the methods indicated previously. Then combine these measurements to give

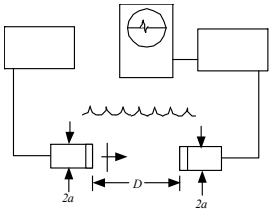
$$s(\omega) = t_R(\omega) t_G(\omega) V_i(\omega)$$

$$s(\omega) = \frac{K Z_o^e S_{vI}^B}{\left(Z_{in}^{B;e} R_{11} + R_{12}\right) + \left(Z_{in}^{B;e} R_{21} + R_{22}\right) Z_o^e} \frac{Z_r^{A;a} S_{vI}^A}{\left(Z_{in}^{A;e} T_{11} + T_{12}\right) + \left(Z_{in}^{A;e} T_{21} + T_{22}\right) Z_i^e} V_i(\omega)$$

Now, consider how we can experimentally measure the system function. There are two methods. The first is to relate the system function to all of the underlying components and measure those components. This is lengthy but possible since we have demonstrated how to measure or model all of those components.

The System Function

B. Measure $V_R(\omega)$, model $t_A(\omega)$, compute $s(\omega) = \dfrac{V_R(\omega)}{t_A(\omega)}$



$$t_A = 2\exp[-\alpha(\omega)D]\left\{ \exp(ik_f D) - \frac{4}{\pi}\int_0^{\pi/2} \sin^2 u \exp\left[ik_f \sqrt{D^2 + 4a^2\cos^2 u}\, du\right]\right\}$$

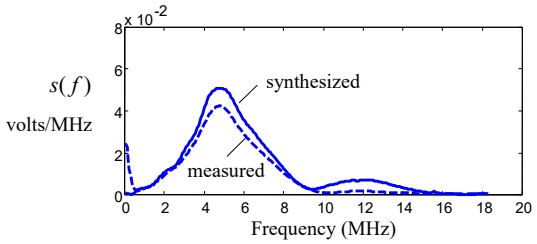
$$s(\omega) = \frac{V_R(\omega)}{t_A(\omega)}$$ deconvolution by straight division

Wiener filter deconvolution $$s(\omega) = \frac{V_R(\omega) t_A^*(\omega)}{|t_A(\omega)|^2 + \varepsilon^2 \max\left\{|t_A(\omega)|^2\right\}}$$

The second method, which is much simpler, is to measure the received voltage in a calibration setup where the acoustic/elastic transfer function can be modeled explicitly and then obtain the system function by deconvolution (division in the frequency domain). Shown is an example pitch-catch setup where the two transducers are placed in an aligned fashion in a water bath and the acoustic/elastic transfer function is given. To stabilize the deconvolution it is not implemented by straight division but through the use of a Wiener filter that includes a small filter constant, ε.

The System Function

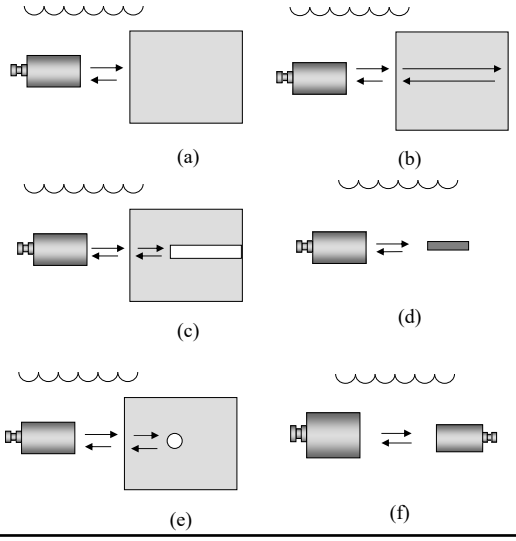
Comparison of system factor obtained with both methods.
method A : synthesized
method B: measured



Here is an example of using both of these methods to obtain the system function. We see that generally we have good agreement.

The System Function

Example setups where t_A is known

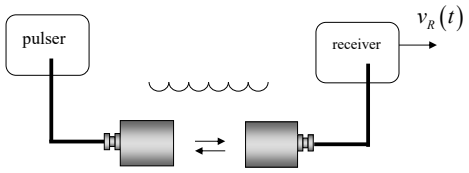


Here are a wider class of setups where the acoustic/elastic transfer function can be modeled explicitly and used to determine the system function through deconvolution. Most of these are for pulse-echo setups, but we do have one setup suitable for a pitch-catch or through-transmission inspection.

The System Function

If all the electrical and electromechanical components of an ultrasonic measurement system are measured to obtain s(ω), then this s(ω) in conjunction with a knowledge of t_A can used to determine the output voltage of a measurement system.

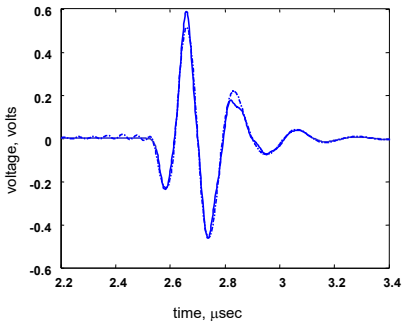
The following examples are shown for this setup:



Although direct measurement of the system function through deconvolution is much simpler than measuring and combining all the underlying system components , a knowledge of those components lets us examine how the measured voltage is affected by all the individual components so that we could, for example, predict how a change of a transducer(s) sensitivity affects the measured voltage when designing a measurement system.

Here, we will give a number of examples of the voltage versus time waveforms predicted for the setup shown compare to the actual measured signals when all the components of the system function are measured.

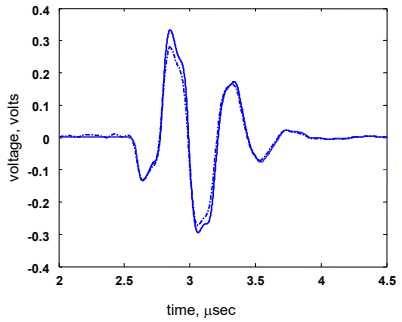
The System Function



Results for a pair of 5MHz, 6.35 mm diameter transducers.
solid line – directly measured output voltage in an ultrasonic pitch-catch setup.
dashed-dotted line – voltage signal synthesized by modeling all the ultrasonic components and performing an inverse FFT to obtain the time domain signal.

First example. All the system components were measured and combined with the acoustic/elastic transfer function to obtain the measured voltage in the frequency domain that was then inverted into the time domain and compared with the directly measured voltage signal.

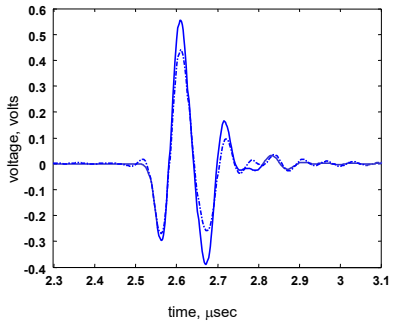
The System Function



Results for a pair of 2.2 5MHz, 12.7 mm diameter transducers.
solid line – directly measured output voltage in an ultrasonic pitch-catch setup.
dashed-dotted line – voltage signal synthesized by modeling all the ultrasonic components and performing an inverse FFT to obtain the time domain signal.

Second example. Again, all the system components were individually measured.

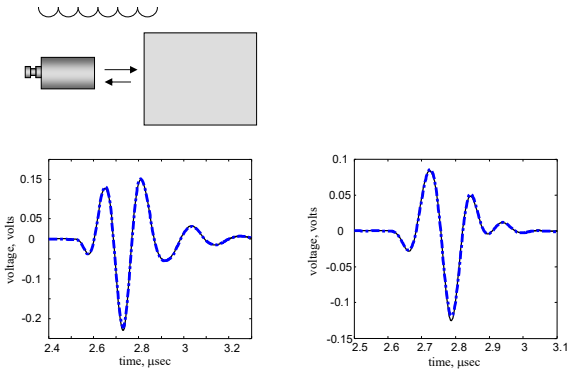
The System Function



Results for a pair of 10 MHz, 6.35 mm diameter transducers.
solid line – directly measured output voltage in an ultrasonic pitch-catch setup.
dashed-dotted line – voltage signal synthesized by modeling all the ultrasonic components and performing an inverse FFT to obtain the time domain signal.

Third example. Again, all the system components were individually measured.

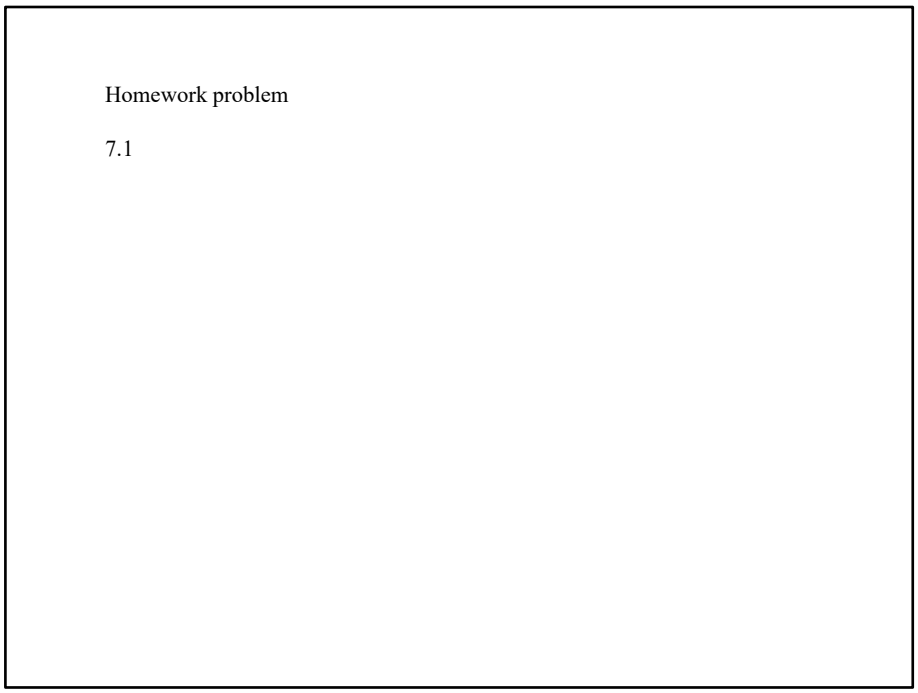
The System Function



(a) (b)

(a) Results for a 5MHz, 6.35 mm diameter transducer.
(b) Results for a 10 MHz, 6.35 mm diameter transducer
solid line – directly measured output voltage in the ultrasonic pulse-echo setup.
dashed-dotted line – voltage signal synthesized by modeling all the ultrasonic components and performing an inverse FFT to obtain the time domain signal.

All the previous examples were for a pitch-catch setup. Here are examples for several transducers in the pulse-echo setup shown above.



Homework problem from Chapter 7.

Transducer Sound Radiation

An important part of the acoustic/elastic transfer function are the waves generated by an ultrasonic transducer. We will describe those waves here.

Learning Objectives

Planar Immersion Transducer
- on-axis near field - far field
- radiation into solid-normal incidence, plane interface
- diffraction correction

Spherically focused transducer
- on axis field
- focal spot size
- radiation into solid-normal incidence, plane interface
- diffraction correction

We will examine with models some of the basic characteristics of the wave fields generated with planar and spherically focused immersion transducers.

Learning Objectives (continued)

Contact P-wave transducer on a solid
- wave types present
- directivity functions

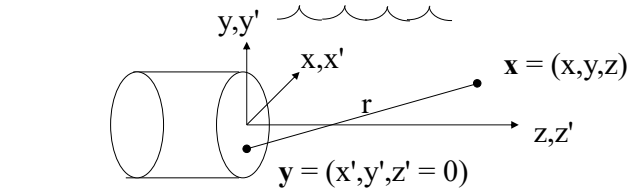
Angle beam shear wave transducer

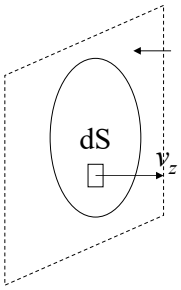
Overview of beam theories

We will also examine the wave fields of other types of transducers such as contact and angle beam transducers and give a brief summary of the types of theories that are available to model transducers.

Ultrasonic Beam Models

Planar transducer radiating into a fluid





Can show that each area element that is in motion acts like a source of spherical waves:

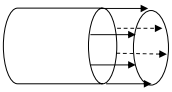
$$dp(x,y,z,\omega) = \frac{-i\omega\rho v_z(x',y',\omega)dS}{2\pi}\frac{\exp(ikr)}{r}$$

First, consider a planar (unfocused) immersion transducer radiating into a fluid. We will model the transducer (in the frequency domain) by assuming that there is a velocity field acting normal to the transducer face over the surface, S, of its face and that the rest of the plane is motionless (called a rigid baffle). If we examine a small element, dS, of that surface, we can show that this element acts like a point source and radiates a spherical wave and generates a small pressure, *dp*, which is shown.

Adding up all such sources over the face of the transducer gives the Rayleigh-Sommerfeld Integral

$$p(x,y,z,\omega) = \frac{-i\omega\rho}{2\pi}\int_S \frac{v_z(x',y',\omega)\exp(ikr)}{r} dS(\mathbf{y})$$

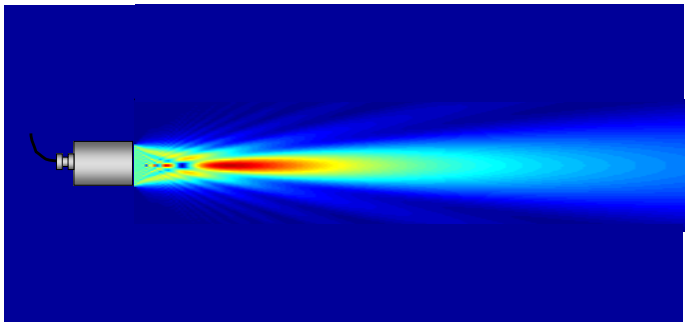
If we let $v_z(x',y',\omega) = v_0(\omega)$ (piston model)



$$\Rightarrow \quad p(x,y,z,\omega) = \frac{-i\omega\rho v_0(\omega)}{2\pi}\int_S \frac{\exp(ikr)}{r} dS(\mathbf{y})$$

If we add up all such spherical waves then the total pressure at any point in the fluid is given by the **Rayleigh-Sommerfeld integral (also called the Rayleigh integral)**. For a piston transducer this integral becomes the simpler form shown. In this form we are simply adding up (in integral form) spherical waves originating the face of the transducer.

plot of the magnitude of the pressure for a 5 MHz, 0.5 inch diameter transducer radiating into water

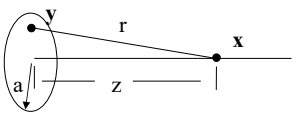


Note: horizontal axis scale is much larger than vertical axis scale

If we numerically evaluate the Rayleigh-Sommerfeld integral for a 5 MHz, 0.5 inch diameter immersion transducer, we can show a cross-sectional plot of the pressure wave field, where red indicates high pressure and blue is low pressure. We see that close to the transducer there is a complex field and that overall the wave field is well collimated, i.e. it is contained mostly in a cylindrical region extending into the fluid from the transducer face. Note that what we are seeing here is the pressure in the frequency domain at a frequency of 5 MHz. A real transducer puts out a pulse of sound but if we compute the FFT of a piston transducer where the velocity on its face is a delta function in time and examine the 5 MHz component of that response at all points in the wave field, we obtain this result.

$$p(\mathbf{x},\omega) = \frac{-i\omega\rho v_0}{2\pi}\int_S \frac{\exp(ikr)}{r}\,dS(\mathbf{y})$$

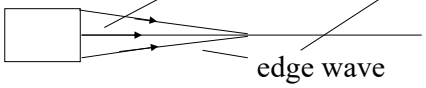
For on-axis response of a circular transducer of radius, a



It can be shown that the area element can be written as dS = r dr dϕ

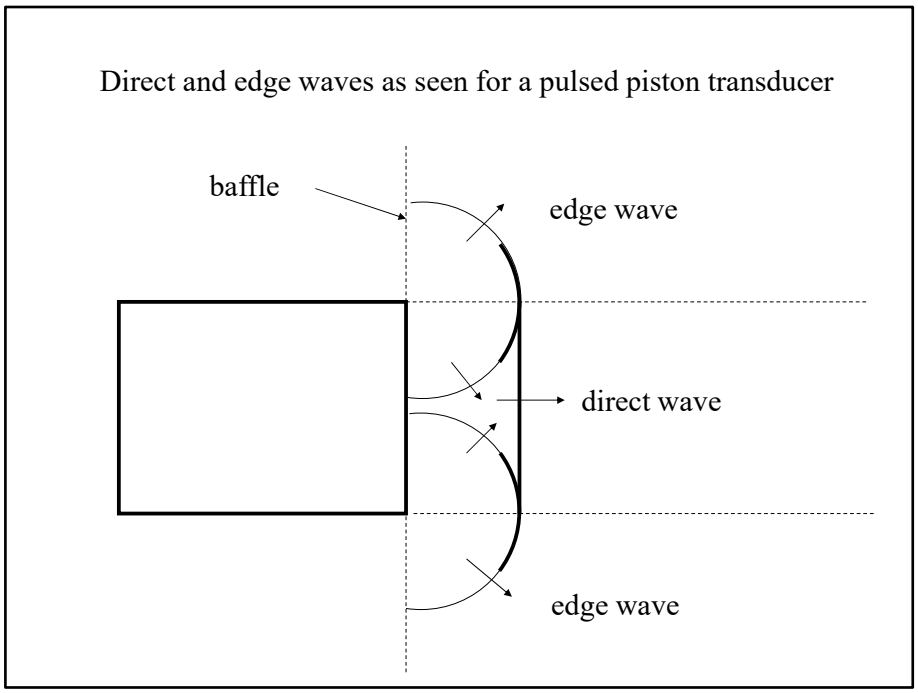
$$p(z,\omega) = \rho c v_0\left[\ \exp(ikz) - \exp\left(ik\sqrt{a^2+z^2}\right)\ \right]$$

direct wave



edge wave

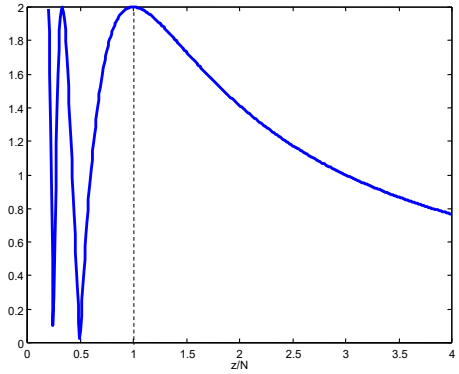
For points in the fluid along the central axis of the transducer we can integrate the Rayleigh-Sommerfeld integral and obtain an explicit expression for the on-axis pressure. That expression has two terms which we can identify as a direct wave from the face of the transducer and a wave that has come from the edge of the transducer face.



If we examine the full wave field of the transducer we can show that the direct wave is actually a plane wave that exists only within a cylinder that extends from the transducer face while the edge wave is in the form of a wave that appears to originate from the edge of the transducer and in 3-d looks like a half “donut” in shape.

on-axis pressure:

$$\frac{|p|}{\rho c v_0}$$



a = 6.35 mm
f= 5 MHz
c = 1500 m/s

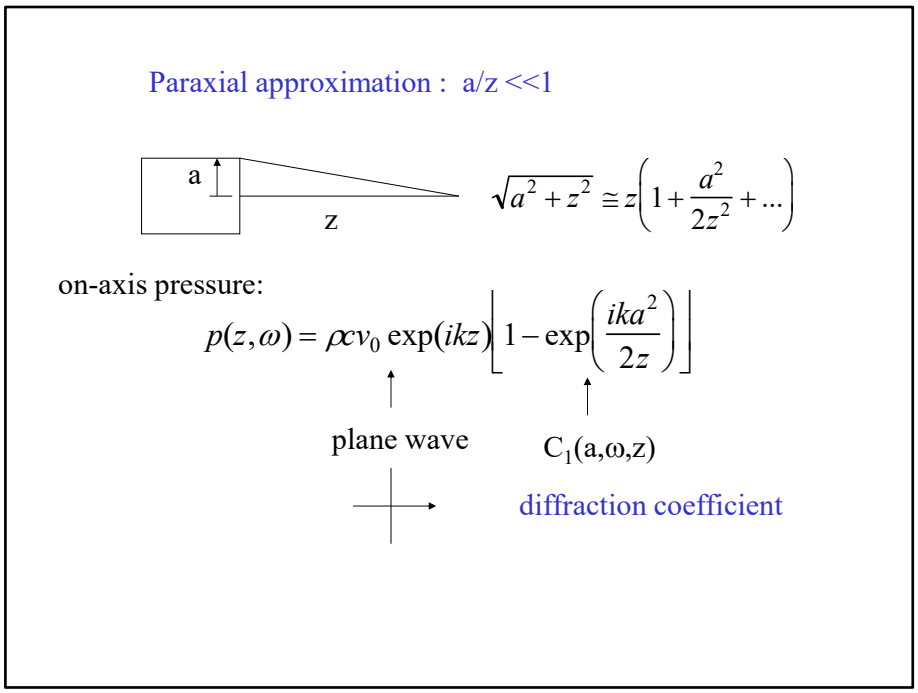
Approximate Near field distance $N = a^2/\lambda$

Maxima: $z = N/(2m+1)$ $m = 0,1,2,\ldots$

Minima: $z = N/2n$ $n = 1,2,3,...$

Example: for a 5 MHz, 1/2 in. diameter transducer radiating into water
N = 5 in. (approx.)

If we plot the magnitude of the on-axis pressure we see that for distance less than a near field distance, N, there are successive maxima and minima in the pressure. This is called the **near field** of the transducer. At about three near field distances or greater the pressure simply decays like 1/z in magnitude, which is the characteristic decay seen for a point source. This makes sense since sufficiently far from the transducer we expect the whole transducer itself will appear as a small point-like source. This region z > 3N is called the **far field** region of the transducer. We will say more about the far field later.

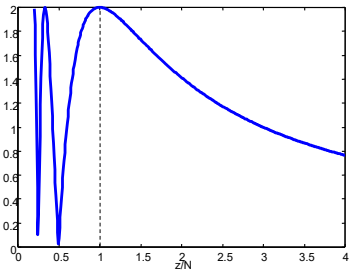


If we assume that we are not too close to the transducer so that we can assume a/z <<1 (note this does NOT assume we are in the far field) then we can use this so-called **paraxial approximation** to write the on-axis pressure as simply a plane wave multiplied by a correction factor called a **diffraction coefficient**. Thus, in the paraxial approximation we can treat the on-axis wave field as a **quasi-plane wave**. Basically, the paraxial approximation assumes the waves are all approximately traveling in the z-direction. From our previous cross-sectional view where we saw the pressure wavefield was well-collimated we can expect the paraxial approximation can be used also for off-axis points, not just the on-axis points seen here.

on-axis pressure:

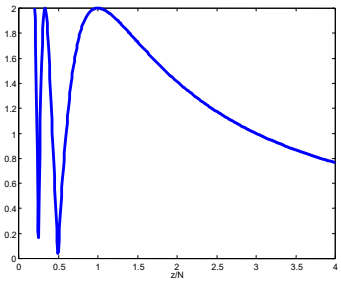
$\frac{|p|}{\rho c v_0}$

exact result



$\frac{|p|}{\rho c v_0}$

paraxial result



Here we can compare the exact on-axis pressure with the paraxial approximation result, showing we capture both the near field and far field behavior well with the paraxial approximation. Generally, the paraxial approximation is valid if we are greater than about a transducer diameter away from its face.

```
function p  = on_axis(zN, A,c,F)
% exact on axis pressure from a piston source
%radiating into a fluid. A is radius in mm, c the
%wavespeed of the fluid in m/sec, F the frequency in MHz,
% zN is the distance in the fluid divided by the near field
%distance a^2/lamba  (lamba is the wavelength)
al= 1000*A*F/c;    % a/lamba
ka = 2*pi*al;              % ka for the transducer
kz = ka*al*zN;
ke = 2*pi*(al^2).*sqrt(zN.^2 + (1/al)^2);
p = exp(i*kz) - exp(i*ke);
```

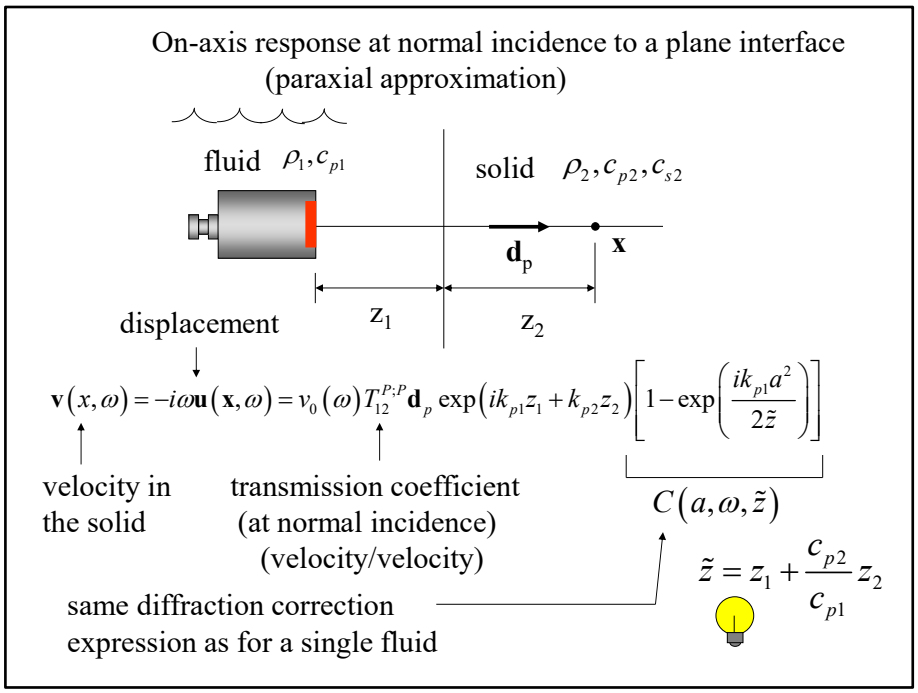
Here is the MATLAB code for calculating the exact on-axis pressure.

```
function p = par_on_axis(zN, A,c,F)
% paraxial axis pressure from a piston source
%radiating into a fluid. A is radius in mm, c the
%wavespeed of the fluid in m/sec, F the frequency in MHz,
% zN is the distance in the fluid divided by the near field
%distance a^2/lamba
al= 1000*A*F/c;    % a/lamba
ka = 2*pi*al;              % ka for the transducer
kz = ka*al*zN;
ke = ka./(2*al.*zN);
p = exp(i*kz).*(1 - exp(i*ke));
```

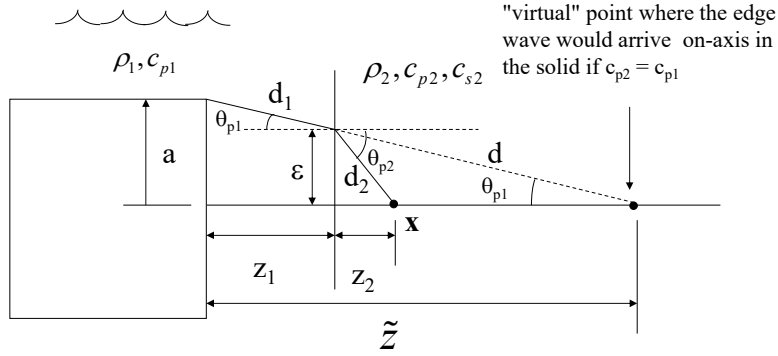
Here is the MATLAB code for calculating the on-axis pressure in the paraxial approximation.

```
MAT> z = linspace(.2, 4,500);
MAT> p = on_axis(z,6.35,1500,5);
MAT> plot(z, abs(p))
MAT> xlabel('z/N')

MAT> p = par_on_axis(z, 6.35, 1500, 5);
MAT> plot(z, abs(p))
MAT> xlabel('z/N')
```

Here is the MATLAB code for generating the on-axis comparison plots we have just seen.



Since the wave field of the transducer behaves like a quasi-plane wave in the paraxial approximation, we can use plane wave theory and obtain the wave field in more complex situations such as where an immersion transducer radiates a compressional wave into a solid, where the transducer is at normal incidence to a plane fluid-solid interface. In this case we can write the velocity field as that of a plane wave that has traveled in the fluid and across the plane interface multiplied by a diffraction correction. In this case we can show that the diffraction coefficient is the same as that for the single fluid problem where the distance z is replaced by the equivalent distance shown.



$$\varepsilon = d_2 \sin\theta_{p2} = d \sin\theta_{p1} \quad \text{so} \quad d = \frac{\sin\theta_{p2}}{\sin\theta_{p1}} d_2 = \frac{c_{p2}}{c_{p1}} d_2$$

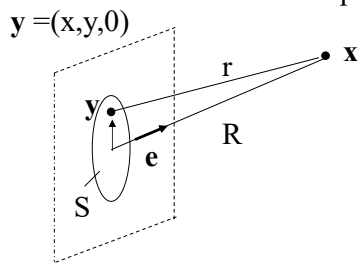
which gives, in the paraxial approximation

$$\frac{a^2}{\tilde{z}} \cong \frac{a^2}{d_1 + d} = \frac{a^2}{d_1 + \frac{c_2}{c_1} d_2} \cong \frac{a^2}{z_1 + \frac{c_2}{c_1} z_1}$$

We can give a physical interpretation of the equivalent distance by examining a wave that arrives on the axis from transducer edge. Using the geometry, Snell's law, and the paraxial approximation, we see that the equivalent distance is just the distance along the axis the edge wave would travel if it had simply been propagating in a single medium.

Far-field beam of a planar piston transducer

The far-field is usually defined as z > 3N - also called the "spherical wave region"



$$r = \sqrt{(\mathbf{x}-\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})}$$
$$= \sqrt{(R\mathbf{e}-\mathbf{y})\cdot(R\mathbf{e}-\mathbf{y})}$$
$$\cong R\sqrt{1-2\mathbf{e}\cdot\mathbf{y}/R}$$
$$\cong R-\mathbf{e}\cdot\mathbf{y}$$

$$p(\mathbf{x},\omega) = \frac{-i\omega\rho v_0}{2\pi}\frac{\exp(ikR)}{R}\int_S \exp(-ik\mathbf{e}\cdot\mathbf{y})\,dxdy$$

Here we will examine the far field region of a piston transducer in more detail. We will see that we can get explicit results for the pressure anywhere, not just on axis. We can examine the radius, r, from a point on the transducer face to a general point in the fluid and approximate that distance in terms of the radius, R, from the center of the transducer and the unit vector, **e**, which defines the direction to the point in the fluid. As a function of R, we see the behavior exp(ikR)/R is just that of a spherical wave so in the far field the wavefield of the transducer behaves like a spherical wave whose angular dependency is governed by the remaining integral . The remaining integral is one we can perform in a number of cases.

Define the 2-D spatial Fourier transform of Θ , where $\Theta = \begin{cases} 1 & in\ S \\ 0 & otherwise \end{cases}$

as

$$F\left(e_x, e_y, \omega\right) = \frac{1}{\left(2\pi\right)^2} \iint_S \exp\left(-ip_x x - ip_y y\right) dxdy \qquad p_x = ke_x$$

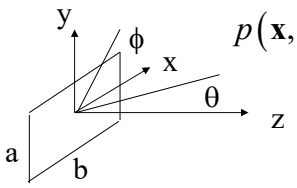
$$= \frac{1}{\left(2\pi\right)^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Theta\left(x.y\right) \exp\left(-ip_x x - ip_y y\right) dxdy \qquad p_y = ke_y$$

Then the far field pressure can be written as

$$p\left(\mathbf{x}, \omega\right) = \underbrace{-2\pi i \omega \rho v_0 F\left(e_x, e_y, \omega\right)}_{\text{angular beam profile}} \underbrace{\frac{\exp\left(ikR\right)}{R}}_{\text{spherical wave}}$$

The remaining integral can be expressed as a 2-D Fourier transform, F, of a characteristic function, Θ , which is just equal to one on the plane of the transducer face and zero outside that face. This integral gives us the **angular beam profile**.

Rectangular Piston Transducer



$$p(\mathbf{x},\omega) = -2\pi i\omega\rho v_0 F \frac{\exp(ikR)}{R}$$

$$F(e_x, e_y, \omega) = \frac{ab}{(2\pi)^2} \frac{\sin\left(\frac{kbe_x}{2}\right)\sin\left(\frac{kae_y}{2}\right)}{\left(\frac{kbe_x}{2}\right)\left(\frac{kae_y}{2}\right)}$$

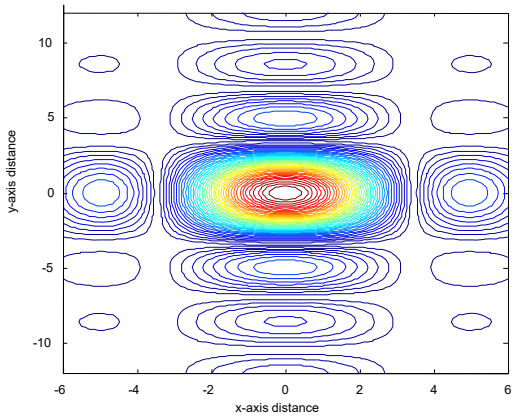
In spherical coordinates

$$e_x = \sin\theta\cos\phi$$
$$e_y = \sin\theta\sin\phi$$

Here is the expression for the angular beam profile of a rectangular transducer in terms of spherical angles.

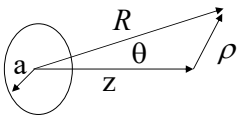
Example far-field pattern of a rectangular transducer

a = 3 mm in y-direction , b = 6 mm in x-direction
R = 70 mm, f = 5 MHz.



If we make a contour plot of the angular beam profile in a plane parallel to the transducer face in the far field we see a central lobe and multiple side lobes

Circular Piston Transducer

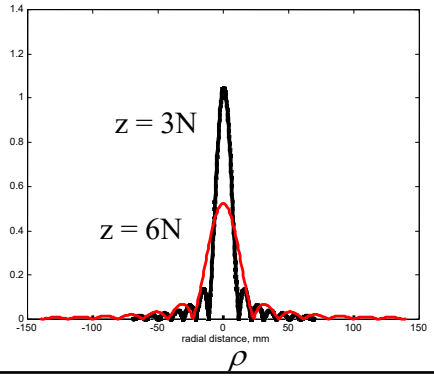


$$p(\mathbf{x},\omega) = -2\pi i\omega\rho v_0 F \frac{\exp(ikR)}{R}$$

$$e_\rho = \sqrt{e_x^2 + e_y^2} = \sin\theta = \frac{\rho}{R}$$

$$F(e_x, e_y, \omega) = \frac{a^2}{2\pi} \frac{J_1(ke_\rho a)}{(ke_\rho a)}$$

$$\frac{|p|}{\rho c v_0}$$



ρ

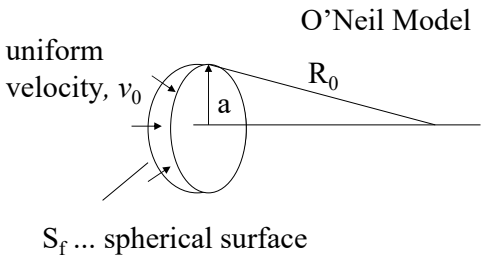
Similarly for a circular piston transducer we can obtain the angular beam profile in terms of a Bessel function. Here, we plot the magnitude of the pressure at both three and six near field distances from the transducer, showing how the main lobe of the transducer and the side lobes spread as we get further from the transducer and , of course, the amplitude gets smaller due to the spherical spreading term

```
function  [p, rho] = far_field(ang,A, c, F, RN)
% far_field computes the normalized far field pressure
% for a circular piston (omitting the exp(ikR) phase term)
% A is the radius of the transducer in mm, c the wavespeed
%in m/sec, F the frequency in MHz, and RN is
%the normalized radial distance in near field units.
% rho is the transverse distance (normal to z)  in mm
ka = 2*pi*(1000*A*F/c);
al= 1000*A*F/c;
x = ka*sin(ang*pi/180);
rho =RN*(A*al)*sin(ang*pi/180);
p = -i*(ka/(al*RN))*besselj(1,x)./(x+eps*(x ==0));
```

```
MAT> ang = linspace(-10, 10,500);
MAT> [p,r] =far_field(ang,6.35,1500,5,3);
MAT> plot(r,abs(p), '--')
MAT> hold on
MAT> [p,r] =far_field(ang,6.35,1500,5,6);
MAT> plot(r,abs(p), 'red')
MAT> xlabel('radial distance, mm')
```

Here is the MATLAB code for obtaining the plots on the previous page.

Spherically Focused Piston Transducer Radiating Into a Fluid

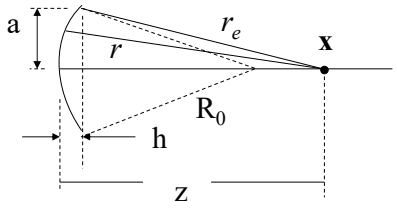


$$p(\mathbf{x},\omega) = \frac{-i\omega\rho v_0}{2\pi}\int_{S_f}\frac{\exp(ikr)}{r}\,dS(\mathbf{y})$$

A real spherically focused transducer is made by placing an acoustic lens on the face of the transducer. However, O'Neil has shown we can also obtain the same focusing by placing a uniform velocity on a spherical surface, which is the model we will use here. Thus, the form of the model is the same as the Rayleigh-Sommerfeld planar transducer model but we now have to integrate over a spherical surface.

For **x** on the central axis

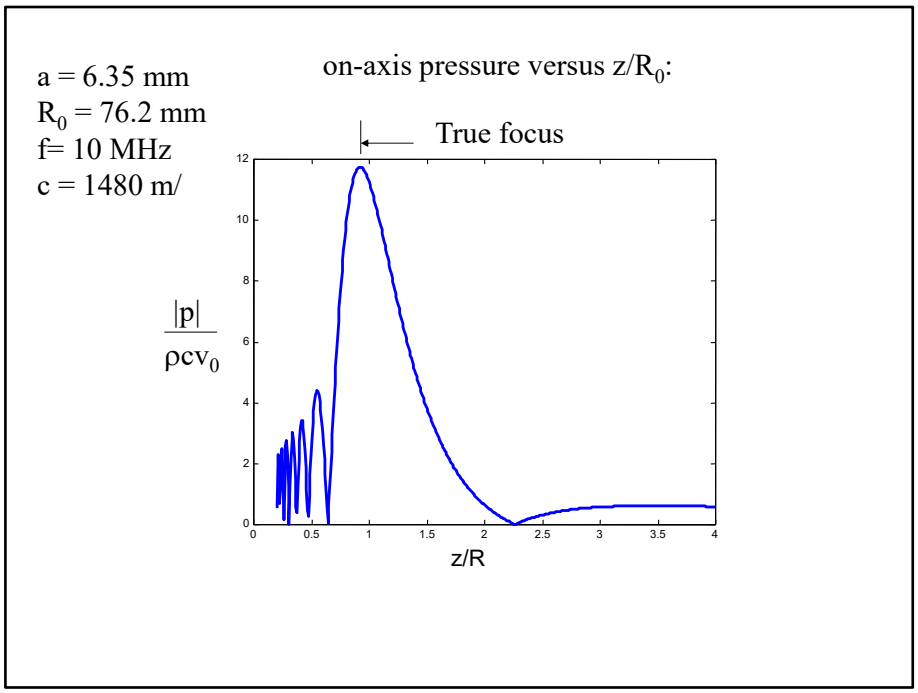
$$dS = r\, dr\, d\phi / q_0 \qquad q_0 = 1 - z/R_0$$



$$p(\mathbf{x},\omega) = \frac{\rho c v_0}{q_0}\left[\exp(ikz) - \exp(ikr_e)\right]$$

$$r_e = \sqrt{(z-h)^2 + a^2} \qquad h = R_0 - \sqrt{R_0^2 - a^2}$$

On the axis of the transducer we can use the O'Neil model and, like the planar transducer case, obtain explicit results for the on-axis pressure.



If we plot the on-axis pressure we see the pressure is enhanced near the **geometrical focus** where z/R =1 but the maximum pressure occurs at a slightly smaller distance called the **true focus**. As the frequency gets smaller the distance grows between the geometrical focus and the true focus. Again, we see rapid oscillations in the near field. However, unlike the planar transducer we can see one or more nulls beyond the maximum pressure location like the one seen here. In some cases, however, such nulls may not exist.

```
function p = focused_on_axis(zR, A,c,F,R)
% on axis pressure of a spherically focused probe
% as a function of the normalized distance, zR = z/R
%A, radius of the transducer in mm. R , focal length in mm.
%c, the wave speed in m/sec, and F the frequency in MHz
al=1000*A*F/c;
ka=2*pi*al;
zN=(R/A)*(1/al)*(zR);
kz=ka*al*zN;
kR=2000*pi*F*R/c;
kh=kR-sqrt(kR^2-ka^2);
kre=sqrt((kz-kh).^2+ka^2);
p = (exp(i*kz) -exp(i*kre))./(1-kz./kR);
```

```
MAT> zR=linspace(.2,4,500);
MAT> p = focused_on_axis(zR,6.35,1480,10,76.2);
MAT> plot(zR,abs(p))
MAT> xlabel('z/R')
```

Here is the MATLAB code for the spherically focused transducer on-axis behavior.

Paraxial Approximation

$$r_e \cong z + \frac{a^2 q_0}{2z} \qquad q_0 = 1 - z/R_0$$

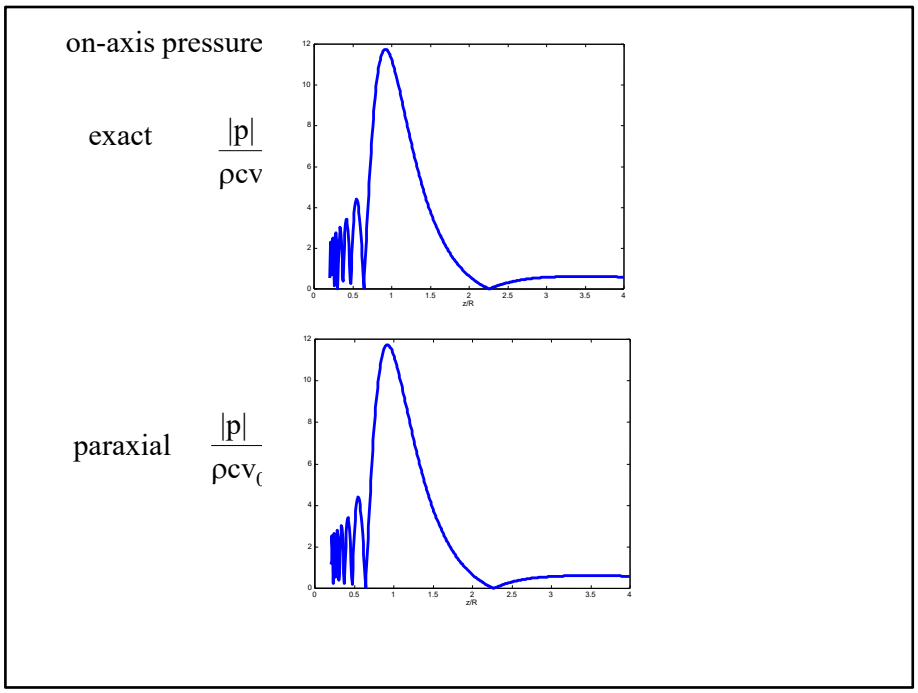
on-axis pressure:

$$p(z,\omega) = \rho c v_0 \exp(ikz)\left\{\frac{1}{q_0}\left[1 - \exp\left(\frac{ika^2 q_0}{2z}\right)\right]\right\}$$

plane wave

diffraction coefficient
$C_1(a, z, R_0, \omega)$

In the paraxial approximation we can again represent the pressure as a plane wave term multiplied by a diffraction coefficient.



Comparing the exact and paraxial results we see again very good agreement. If the transducer is very tightly focused the paraxial approximation will not be accurate but most NDE transducers are not that tightly focused.

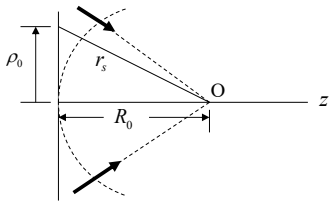
```
function p = par_focused_on_axis(zR, A,c,F,R)
% on axis pressure of a spherically focused probe,paraxial approx.
% as a function of the normalized distance, zR = z/R
%A, radius of the transducer in mm. R , focal length in mm.
%c, the wave speed in m/sec, and F the frequency in MHz
al=1000*A*F/c;
ka=2*pi*al;
zN=(R/A)*(1/al)*(zR);
kz=ka*al*zN;
kR=2000*pi*F*R/c;
qo=1-kz./kR;
p = (1-exp(i*ka*(A/R)*qo./(2*zR))).*qo;
```

```
MAT> zR=linspace(.2,4,500);
MAT> p = par_focused_on_axis(zR,6.35,1480,10,76.2);
MAT> plot(zR, abs(p))
MAT> xlabel('z/R')
```

Here is the MATLAB code for the paraxial on-axis pressure

Another way to model focusing (in the paraxial approximation)

suppose on a planar aperture we have a spherical wave propagating (generated by a lens, for example)



then on the aperture we have a phase given approximately in the paraxial approximation $(\rho_0 / R_0 << 1)$ by

$$\exp\left(-ik\left[r_s - R_0\right]\right) = \exp\left[-ik\left[\sqrt{\rho_0^2 + R_0^2} - R_0\right]\right]$$
$$\cong \exp\left(-ik\rho_0^2 / 2R_0\right)$$

Instead of using the O'Neil model we can model the effects of spherical focusing by considering an inward traveling spherical wave (as might be generated by a lens, for example) that is incident on a planar aperture. In the paraxial approximation we can examine the phase of this spherical wave.

Thus, suppose we use a Rayleigh-Sommerfeld model for a planar transducer and place this phase (in the paraxial approximation) in the integral:

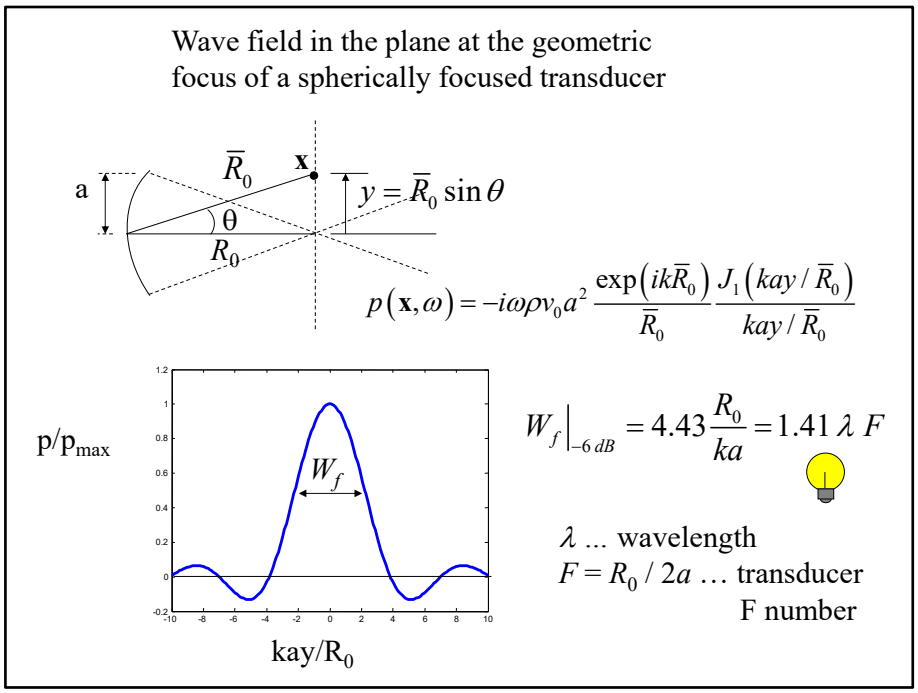
$$p(\mathbf{x},\omega) = \frac{-i\omega\rho\, v_0(\omega)}{2\pi} \iint_S \exp\left(-ik\rho_0^2 / 2R_0\right) \frac{\exp(ikr)}{r} dS$$

Using the paraxial approximation and evaluating this integral exactly for **x** on the transducer axis gives for a circular transducer of radius a:

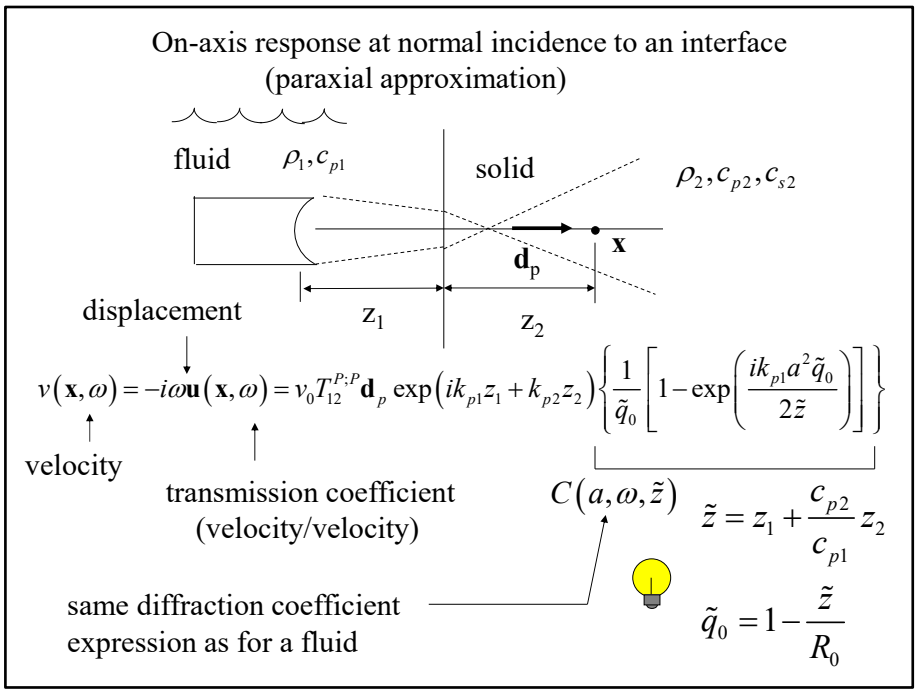
$$p(z,\omega) = \frac{\rho c v_0 \exp(ikz)}{q_0} \left[1 - \exp\left(ika^2 q_0 / 2z\right)\right]$$

Similarly, off-axis values will also represent those from a focused transducer

If we now insert that phase into the Rayleigh-Sommerfeld integral and perform the integrations in the paraxial approximation, we obtain an on-axis result which agrees with the O'Neil model in the paraxial approximation. Other points in the wave field will also agree. This method of generating focusing we will see later is very useful when we discuss a Gaussian beam model of the transducer wave field.

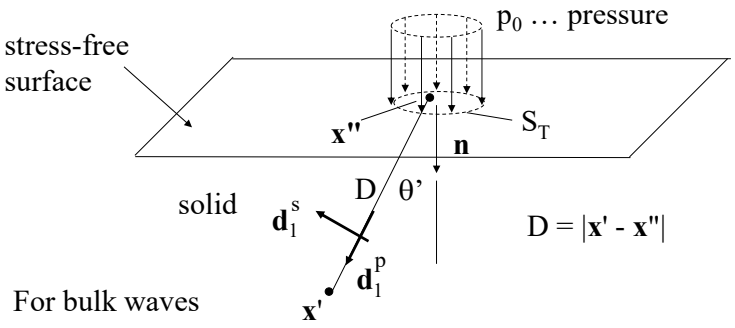


Another case where we can evaluate the O'Neil model exactly is to obtain the wave field in a plane located at the geometric focus of the spherically focused transducer. We see that this field is controlled by a Bessel function behavior that is identical in form with that for the far field behavior of a circular piston transducer. An important measure of the wave field on this plane is the **-6 dB beam width**, W_f , where the peak amplitude has dropped to one half of its on-axis maximum value. We can show that this beam width is controlled by the wave length and the **transducer F-number**.



In the paraxial approximation we can also obtain an explicit expression for the on-axis velocity of a spherically focused P-wave transducer radiating at normal incidence into a solid through a plane interface. We simply multiply the result for a plane wave being transmitted through the interface by a diffraction coefficient which is the same as for the radiation into a fluid, but where the on-axis distance in the fluid case is replaced by an effective distance and the q_0 focusing term also contains that effective distance.

Contact P-wave Transducer Model



$$\mathbf{u}(\mathbf{x}',\omega) = \frac{p_0}{2\pi\rho_1 c_{s1}^2}\int_{S_T} K_s(\theta')\mathbf{d}_1^s \frac{\exp(ik_{s1}D)}{D} dS(\mathbf{x}'')$$

$$+\frac{p_0}{2\pi\rho_1 c_{p1}^2}\int_{S_T} K_p(\theta')\,\mathbf{d}_1^p \frac{\exp(ik_{p1}D)}{D} dS(\mathbf{x}'')$$

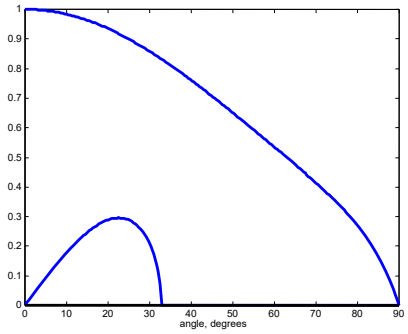
Now, consider a contact P-wave transducer sitting on the plane stress-free surface of an elastic solid. We cannot model this transducer as a piston source since the stiffness of the solid is likely the same order or larger than the stiffness of the transducer. However, since there is a thin fluid layer between the transducer and the solid, it makes sense to model the transducer instead as producing a uniform pressure over its active surface. We will see shortly that such a pressure will produce a complex set of waves, but an important set of those waves consist of radiating spherical P- and S- bulk waves that are contained in Rayleigh-Sommerfeld-like integrals, as shown above. Unlike the immersion transducer case, however, these integrals include directivity factors and polarization vectors.

Directivity functions $K_p(\theta') = \dfrac{\cos\theta' \kappa_1^2 \left(\kappa_1^2/2 - \sin^2\theta'\right)}{2G(\sin\theta')}$

$K_s(\theta') = \dfrac{\kappa_1^3 \cos\theta' \sin\theta' \sqrt{1-\kappa_1^2 \sin^2\theta'}}{2G(\sin\theta')}$

$G(x) = \left(x^2 - \kappa_1^2/2\right)^2 + x^2\sqrt{1-x^2}\sqrt{\kappa_1^2 - x^2}$ $\qquad \kappa_1 = \dfrac{c_{p1}}{c_{s1}}$

K_p , K_s



Here the directivity factors are plotted as a function of the angle from the normal to the surface.

```
function [kp,ks] = directivity(ang, cp, cs)
% computes the directivity functions for a p-wave contact
%transducer. ang is angle in degrees, cp, cs are p- and s-wave
%speeds
k = cp/cs;
angr = ang*pi/180;
x = sin(angr);
c =cos(angr);
g=(x.^2 -k^2/2).^2 + x.^2.*sqrt(1 - x.^2).*sqrt(k^2 - x.^2);
kp = c.*(k^2).*(k.^2/2 -x.^2)./(2.*g);
ks = (k*x <1).*c.*(k^3).*x.*sqrt(1 - k^2.*x.^2)./(2.*g);

MAT> x = linspace(0,90,200);
MAT> [kp,ks] = directivity(x, 5900, 3200);
MAT> plot(x, kp)
MAT> hold on
MAT> plot(x, ks)
MAT> xlabel('angle, degrees')
```

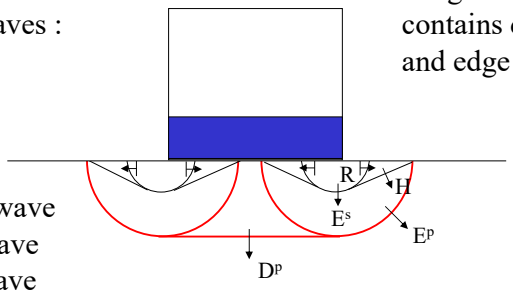
The MATLAB code for the directivities.

For θ' small $K_p = 1$, $K_s = 0$

$$\mathbf{u}(\mathbf{x}',\omega) = \frac{p_0 \mathbf{n}}{2\pi\rho_1 c_{p1}^2} \int_{S_T} \frac{\exp(ik_{p1}D)}{D} dS$$

integral contains direct and edge P-waves

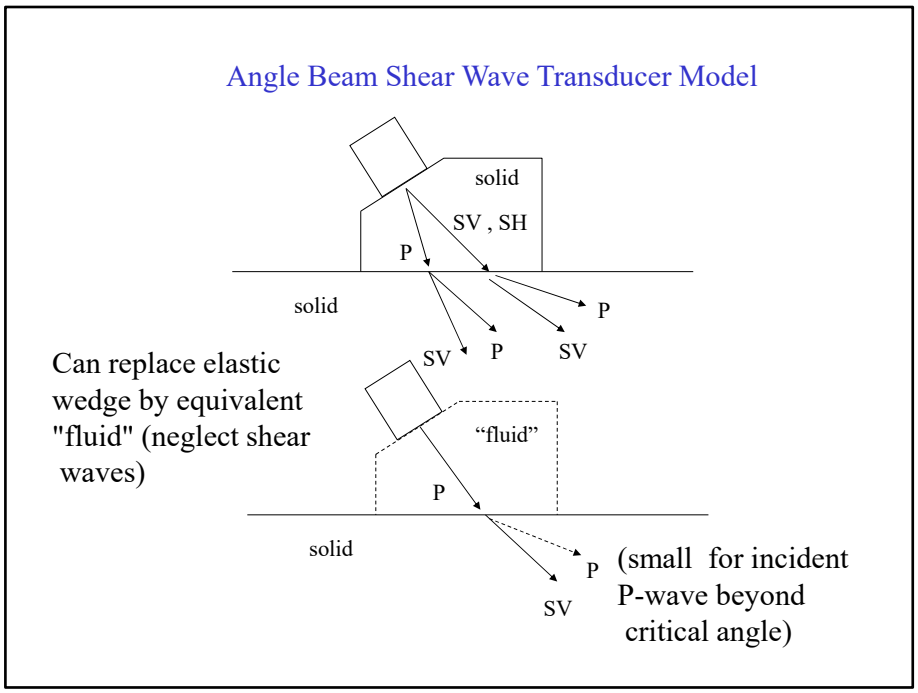
Full set of waves :



Dᵖ … Direct P-wave
Eᵖ … Edge P-wave
Eˢ … Edge S-wave
H … Head wave
R … Rayleigh wave

Like the immersion case, a contact NDE transducer will operate at high frequencies where the transducer beam will be well-collimated and so the fields will be small when the angle from the transducer normal is not small. Thus, for small angles the S-wave directivity will be near zero and the P-wave directivity will be close to one, giving an expression for the displacement in the solid that is again in the form of a Rayleigh-Sommerfeld integral.

We have also shown above a more complete picture of all the waves generated by the contact transducer. We see a direct P-wave and edge P-waves seen previously in the immersion case (red lines). These waves are contained in the Rayleigh-Sommerfeld integral shown above. However, there are also edge S-waves generated and, when the edge P-waves graze along the free surface, additional "head" waves (also called "von Schmidt waves") are generated that link the edge P- and S-waves. Finally, there are also Rayleigh waves that propagate primarily along and near the surface. Thus, the total wave field is rather complex. However, not too close to the transducer and near the transducer central axis, all the waves except the direct and edge P-waves will be small and the Rayleigh-Sommerfeld integral given above will model the wave response.



If a contact transducer is placed on the surface of a wedge at an angle and the wedge is then placed on the surface of an elastic solid (usually with a thin fluid layer between the wedge and the solid) a complex set of transmitted P- and S-waves will be generated (the reflected waves are not shown since the wedge is usually designed to suppress their contributions). If the angle of the wedge is such that the direction of the P-wave in the wedge is above the first critical angle, then primarily an SV-wave will be generated in the solid, producing what is called an **angle beam shear wave transducer** (a small P-wave is present but it can normally be ignored). Since the waves present in this case are the same as those present for a fluid-solid interface, we can replace this configuration by an equivalent fluid-solid configuration and use an immersion transducer model, where we replace the transmission coefficient for a fluid-solid interface by the transmission coefficient for two solids in smooth (shear-stress -free) contact.

Ultrasonic Beam Models

Numerically Intense Models
- EFIT - K. Langenberg
- Finite Elements – W. Lord
- Boundary Elements – F. Rizzo
- Edge Elements – L. Schmerr, T. Lerch

Surface Integral Models
- Generalized Point Source – M. Spies
- Rayleigh- Sommerfeld + High Freq. Asymptotics - L. Schmerr, A. Lhemery, others

Line Integral Models
- Boundary Diffraction Wave – L. Schmerr, T. Lerch

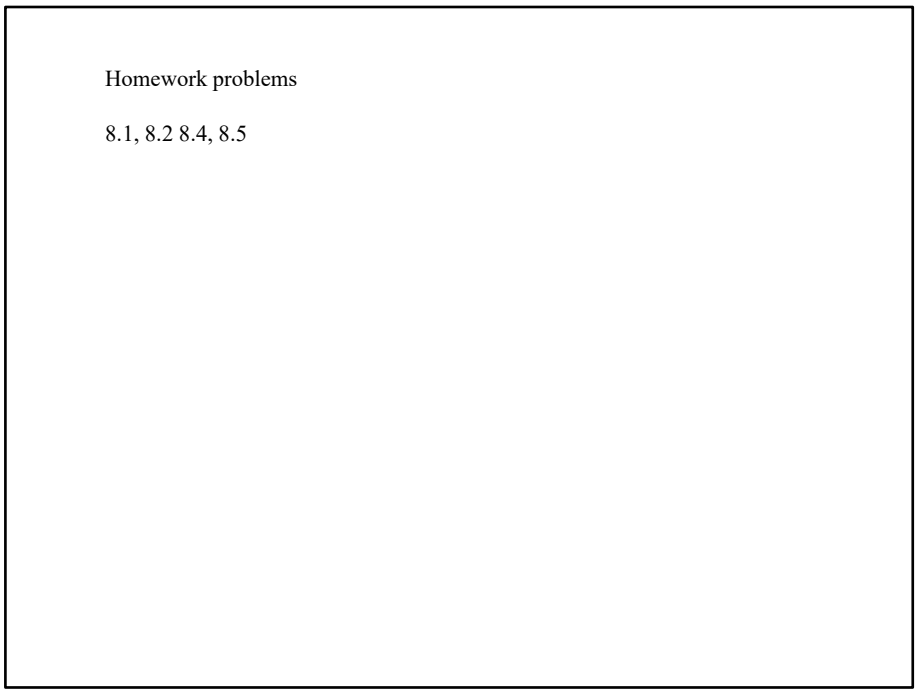
Our discussions in this section have centered around Rayleigh-Sommerfeld integral models of the sound beam generated by a transducer but there are many other transducer models available. There are numerically intense models representing the transducer surface (and in some cases the surrounding media) as a collection of very small elements. There are the surface integral models that superimpose spherical waves from point sources over the transducer face, and there are line integral models that superimpose a plane wave and boundary diffraction waves acting over the edge of the transducer.

Ultrasonic Beam Models

Other Basis Function Models
Gauss-Hermite Models – B. Thompson, T. Gray, B. Newberry, A. Minachi, F. Margetan

Multi- Gaussian Models
A. Minachi, M. Spies, L. Schmerr and M. Rudolph, Cerveny (Seismology)

One can superimpose waves other than spherical waves to represent the transducer wave field such as Gauss-Hermite waves. However, one of the most effective models uses the paraxial approximation and the superposition of a small number of Gaussian beams. We will say more about a **multi-Gaussian beam model** later.



Homework problems from Chapter 8.

Ultrasonic Beam Models

A few references – mostly paraxial models

Lerch, T.P., Schmerr, L.W. and A. Sedov," Ultrasonic beam models: an edge element approach," J. Acoust. Soc. Am., 104, 1256-1265, 1998.

Thompson, R. B. and E.F. Lopez," The effects of focusing and refraction on Gaussian ultrasonic beams," J. Nondestr. Eval., 4, 107-123, 1984.

Newberry, B.P. and R.B. Thompson," A paraxial theory for the propagation of ultrasonic beams in anisotropic solids," J. Acoust. Soc. Am., 85, 2290-2300, 1989.

Schmerr, L.W., Rudolph, M., and A. Sedov," Modeling ultrasonic transducer wave fields for general complex geometries and anisotropic materials," **Review of Progress in Quantitative Nondestructive Evaluation**, D. O. Thompson and D.E. Chimenti, Eds., Plenum Press, New York, 19A, 953-960, 2000.

Here are some references.

Ultrasonic Beam Models

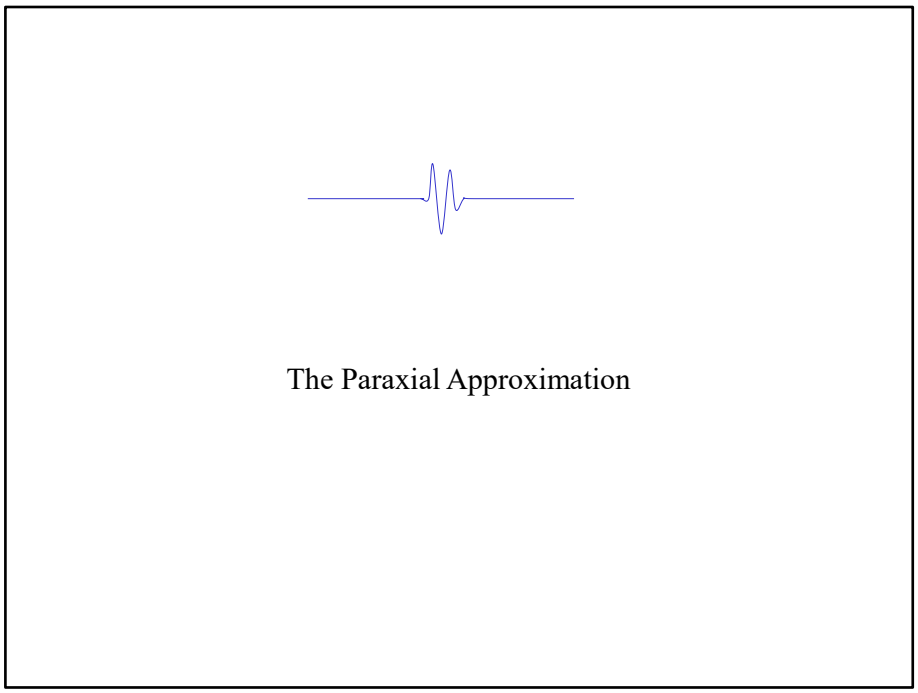
Schmerr, L. W., **Fundamentals of Ultrasonic Nondestructive Evaluation – A Modeling Approach**, **2nd Ed**. ,Springer, Switzerland, 2016.

Spies, M., and M. Kroning," Ultrasonic inspection of inhomogeneous welds simulated by Gaussian beam superposition," **Review of Progress in Quantitative Nondestructive Evaluation**, D. O. Thompson and D.E. Chimenti, Eds., Plenum Press, New York, 18A, 1107-1114, 1999.

Minachi, A., Margetan, F.J., and R.B. Thompson," Reconstruction of a piston transducer beam using multi-Gaussian beams (MGB) and its applications," **Review of Progress in Quantitative Nondestructive Evaluation**, D. O. Thompson and D.E. Chimenti, Eds., Plenum Press, New York, 17A, 907-914, 1989.

Gengembre,N. and A Lhemery," Calculation of wide band ultrasonic fields radiated by water-coupled transducers into heterogeneous media," **Review of Progress in Quantitative Nondestructive Evaluation**, D. O. Thompson and D.E. Chimenti, Eds., Plenum Press, New York, 18A, 1107-1131, 1999.

More references. There are many, many more than listed here.



This will be a brief discussion and overview of the paraxial approximation.

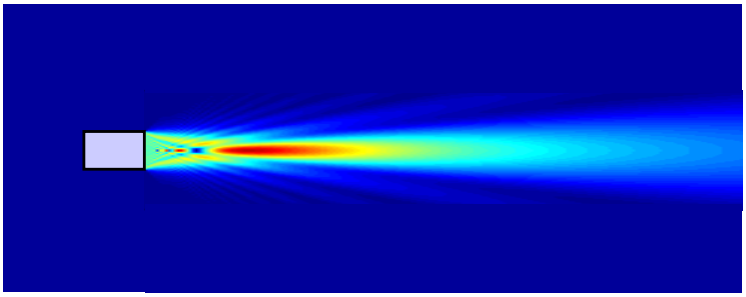
Learning Objectives

Introduction to the paraxial approximation and its importance in beam modeling

Most NDE transducers produce well-collimated beams that can be modeled effectively as the superposition of waves where the paraxial approximation is valid so we want to examine that approximation in more detail. This approximation is also key to developing a multi-Gaussian beam model which is one of the most effective models available for modeling transducer wave fields so the paraxial approximation is a key approximation for that reason as well.

The Paraxial Approximation

A transducer generates a complex beam of sound. Modeling that complexity is a challenging task.

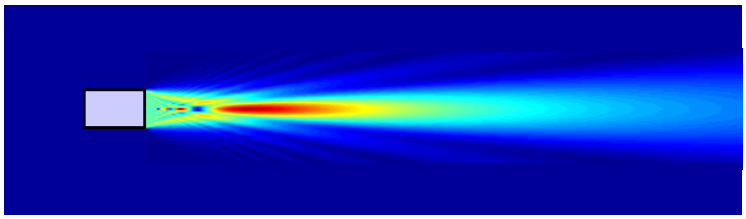


The modeling computational burden can be reduced considerably by introducing approximations. The paraxial approximation is one of the most useful of those approximations.

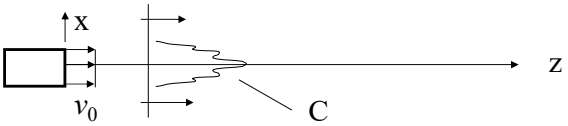
As we have seen a transducer generates a complex sound beam so that modeling that complexity is challenging. The paraxial approximation will help minimize the computational burden immensely.

The Paraxial Approximation

The paraxial approximation models the transducer beam as a quasi-plane wave where most of the sound is propagating in



a given direction with an amplitude profile described by a diffraction coefficient term, C(x,y,z,ω).

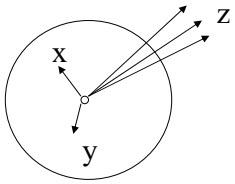


$$p = \rho c v_0 C(x, y, z, \omega) \exp(ikz - i\omega t)$$

In the paraxial approximation the transducer wavefield can be modeled as a plane-wave-like term modified by a diffraction coefficient that accounts for much of the beam complexity. We have seen examples of such diffraction coefficients previously. We can call a wave in this form as a **quasi-plane wave**.

The Paraxial Approximation

To illustrate the paraxial approximation in a simple setting consider a spherical wave. Suppose that we are only interested in the spherical wave field in the vicinity of a particular direction which we will take as the z-axis (i.e. x,y <<z)



$$p = p_0 \frac{r_0 \exp(ikr)}{r}$$

$$v_z \cong v_r \cong \frac{p_0}{\rho c} \frac{r_0 \exp(ikr)}{r}$$

A spherical wave obviously is not well collimated since it travels in all directions. But suppose we consider that wave only near a particular direction which we will call the z-axis here. Shown are the expressions for the pressure and velocity in such a spherical wave traveling in a fluid at high frequencies.

The Paraxial Approximation

Then, since

$$r = \sqrt{x^2 + y^2 + z^2}$$

$$= z\sqrt{1 + \frac{x^2}{z^2} + \frac{y^2}{z^2}}$$

$$\cong z + \frac{\rho^2}{2z}$$

where $\rho = \sqrt{x^2 + y^2}$

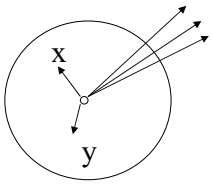
we find

$$p = p_0 \frac{r_0 \exp\left(\frac{ik\rho^2}{2z}\right)}{z} \exp(ikz) = p_0 C(x, y, z; \omega) \exp(ikz)$$

$$v_z = \frac{p_0}{\rho c} \frac{r_0 \exp\left(\frac{ik\rho^2}{2z}\right)}{z} \exp(ikz) = \frac{p_0}{\rho c} C(x, y, z; \omega) \exp(ikz)$$

If we expand the spherical radius r in the neighborhood of this z-axis then we can obtain the pressure and velocity in a quasi-plane wave form.

The Paraxial Approximation



$$p = p_0 \frac{r_0 \exp\left(\frac{ik\rho^2}{2z}\right)}{z} \exp(ikz) = p_0 C(x, y, z; \omega) \exp(ikz)$$

$$v_z = \frac{p_0}{\rho c} \frac{r_0 \exp\left(\frac{ik\rho^2}{2z}\right)}{z} \exp(ikz) = \frac{p_0}{\rho c} C(x, y, z; \omega) \exp(ikz)$$

Note that in obtaining this diffraction correction term we only approximated the amplitude part of the spherical wave to first order (r ~ z) while we approximated the phase to second order. This is because terms neglected in the phase must not only be small with respect to the terms retained but also must be small with respect to 2π if they are to be negligible.

In making this approximation we must treat the phase terms more carefully than the amplitude terms.

The Paraxial Approximation

Consider the wave equation

$$\frac{\partial^2 p}{\partial x^2}+\frac{\partial^2 p}{\partial y^2}+\frac{\partial^2 p}{\partial z^2}-\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}=0$$

Let $p = P(x,y,z,\omega)\exp\left[ikz - i\omega t\right]$

Then $$\frac{\partial^2 P}{\partial x^2}+\frac{\partial^2 P}{\partial y^2}+\frac{\partial^2 P}{\partial z^2}+2ik\frac{\partial P}{\partial z}=0$$

If we assume $$\left|\frac{\partial^2 P}{\partial z^2}\right| << \left|2ik\frac{\partial P}{\partial z}\right|, \left|\frac{\partial^2 P}{\partial x^2}\right|, \left|\frac{\partial^2 P}{\partial y^2}\right|$$ (paraxial approximation)

$$\Rightarrow \quad \frac{\partial^2 P}{\partial x^2}+\frac{\partial^2 P}{\partial y^2}+2ik\frac{\partial P}{\partial z}=0$$

Now, consider the propagation of waves in more general terms. Consider the wave equation for the pressure and write that pressure in a quasi-plane wave form. The paraxial approximation assumes that the second derivative term along an axis (taken here as the z-axis) near which the wave is propagating is much smaller than all the other terms. This is a purely mathematical definition of the paraxial approximation but shortly we will describe the meaning of this assumption in more physical terms.

$$\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2} + 2ik\frac{\partial P}{\partial z} = 0$$

paraxial wave equation

Our paraxial approximation for a spherical wave satisfies this paraxial equation exactly:

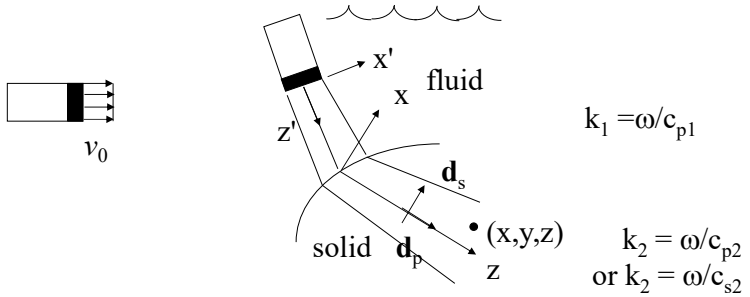
$$P = \frac{p_0 r_0}{z}\exp\left(\frac{ik\rho^2}{2z}\right)$$

There are other solutions, such as Gaussian waves that also satisfy this equation and form important building blocks for modeling ultrasonic transducer radiation in complex problems

Under this mathematical condition the wave equation reduces to the paraxial wave equation for the amplitude of the quasi-plane wave. Our paraxial approximation for the spherical wave satisfies this paraxial wave equation exactly and we will see other solutions as well. The most important of these solutions are Gaussian beams.

The Paraxial Approximation

The paraxial approximation allows one to obtain diffraction coefficient terms for many practical testing setups



$$\mathbf{v}(x,y,z,z',\omega) = v_0 T_{12} \mathbf{d}\, C(x,y,z,z',\omega) \exp(ik_1 z' + ik_2 z)$$

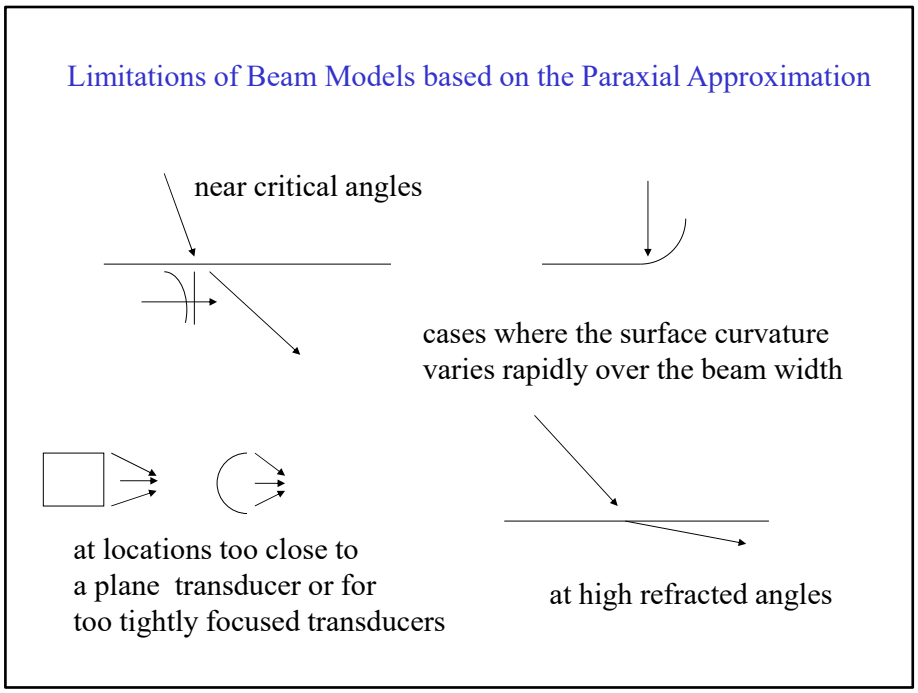
velocity

plane wave transmission coefficient

polarization $\mathbf{d}_p$ or $\mathbf{d}_s$

diffraction coefficient

The paraxial approximation allows us to obtain the diffraction coefficient explicitly in many important testing setups from a number of more exact models so it can reduce the computational burden of predicting transducer wave fields considerably.



The paraxial approximation is, however, as its name implies, an approximation! Thus, it can be inaccurate in some testing situations. Near critical angles at interfaces, for example, the fields are rapidly varying and this variation will render the paraxial approximation (which assumes that basically everything is relatively slowly varying about a single direction) invalid. Similarly, when inspecting through a curved interface where the surface curvature varies rapidly, the paraxial approximation is in error. If we are too close to a planar transducer or are using a very tightly focused transducer the paraxial approximation loses validity since the waves in those cases are traveling at a wide range of angles, not primarily along a single direction as the paraxial approximation assumes. Finally, the rapidly changing wave field when a refracted wave is near to grazing an interface will cause the paraxial approximation to be inaccurate. Fortunately, these cases are not encountered often in most testing situations so that the paraxial approximation is very useful and accurate most of the time.

Gaussian Beams and Transducer Modeling

This will be an overview of Gaussian beams and their use in transducer modeling.

Learning Objectives

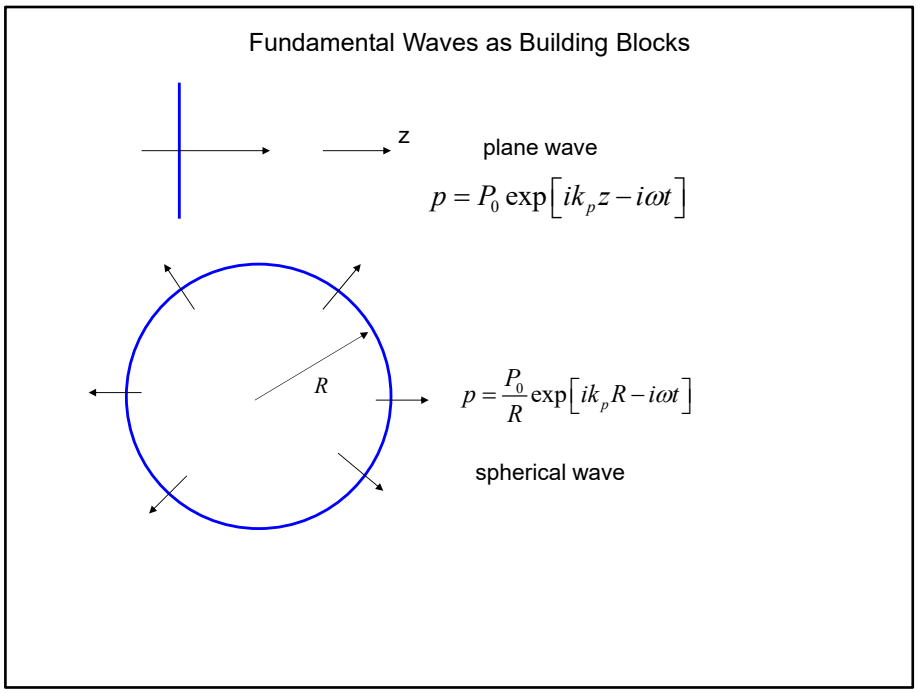
Characteristics of Gaussian Beams

Propagation Laws
Transmission/Reflection Laws

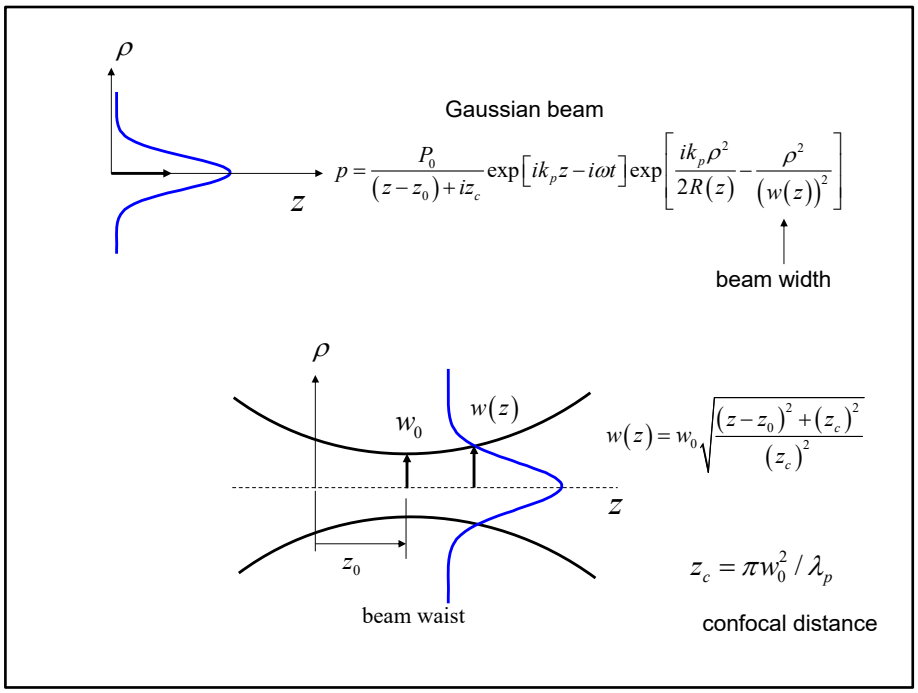
Multi-Gaussian Beam Models for Ultrasonic Transducers

MATLAB Examples

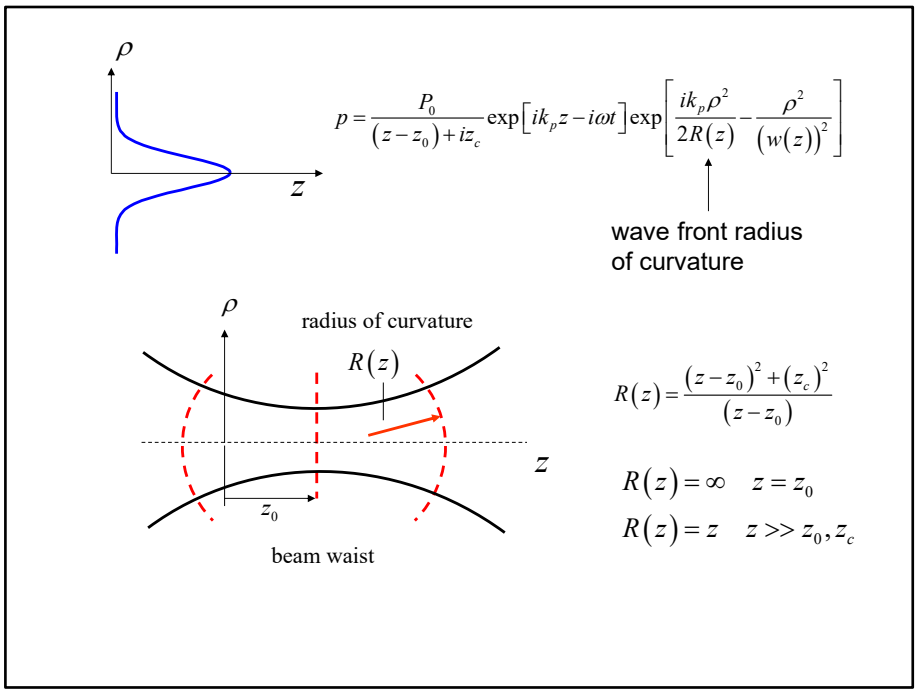
We will discuss some of the important properties of Gaussian beams and show how we can superimpose a small number of such beams to model a transducer wave field. We will also give a number of MATLAB examples of such a **multi-Gaussian beam model**.



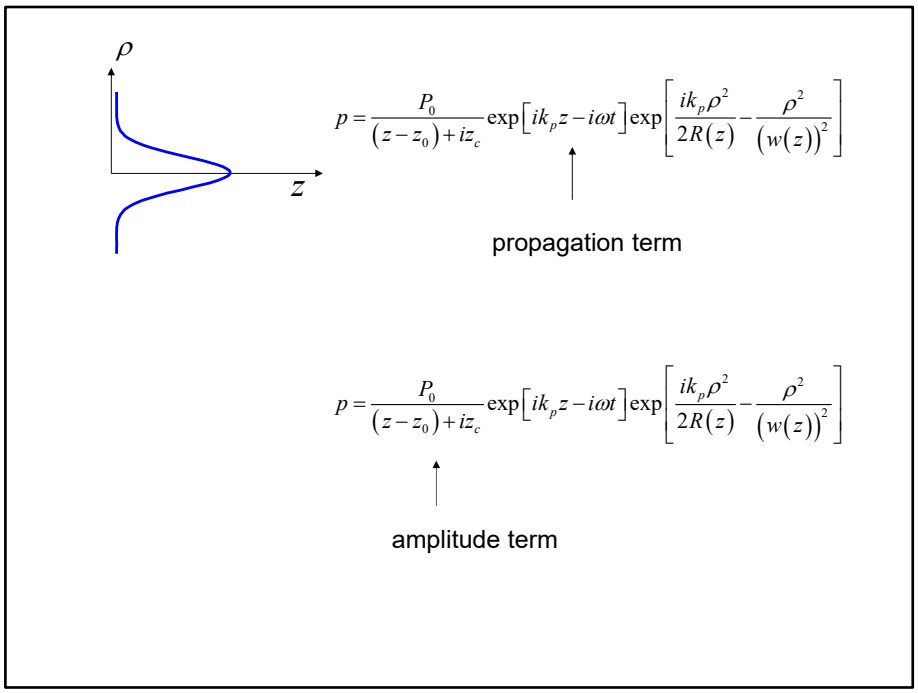
We have already talked about two types of waves, plane waves and spherical waves, that can be used as fundamental building blocks for generating more complex wave fields.



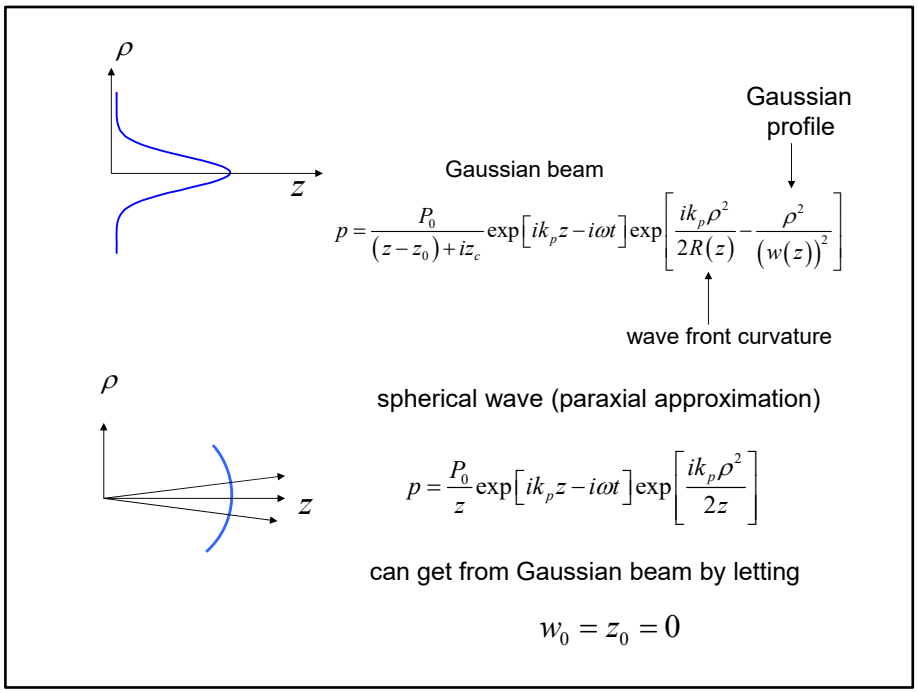
A Gaussian beam is another fundamental building block. Here we will write out the Gaussian beam in a form that is rather complex looking but will be useful to discuss its properties. Later, we will express the Gaussian beam in more compact form. The Gaussian beam is a solution to the paraxial equation. We see that it has an amplitude term which is of the form of a Gaussian function and involves the **beam width**, which is a function of the distance z, the location of the beam waist where the beam width is w_0 , and a distance parameter, z_c , called the **confocal distance**.



There is also a term in the exponential which is a function of R(z), the **radius of curvature of the wavefront**. We see that for distances z greater than the beam waist location this radius is positive, representing an outward spreading wave. At very large z the radius is just the distance z, so it acts like a spherically spreading wave in the paraxial approximation. For distances z less than the beam waist location the radius is negative, representing an inwardly curving wave front. At the beam waist the radius of curvature is infinite so the wavefront is planar. Thus the Gaussian beam can represent many types of propagating waves.

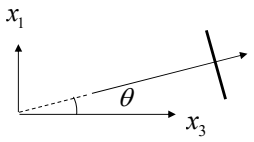


We see that the Gaussian is propagating in the z-direction so it has the familiar propagation term we see for waves. There also is an amplitude term which is controlled by the distance z, the location of the beam waist, and the confocal parameter. For large z this is just the 1/z amplitude term we see for a spherical wave in the paraxial approximation.



If we compare the Gaussian beam with a spherical wave (in the paraxial approximation) we see there are many similarities. In fact, when the width at the beam waist and the beam waist location are both zero, the Gaussian beam coincides with the spherical wave expression.

Paraxial Approximation



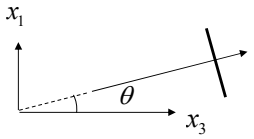
plane wave $p = A\exp\left(ik_p x_1 \sin\theta + ik_p x_3 \cos\theta - i\omega t\right)$

write as $p = P\left(x_1, x_3, \omega\right)\exp\left(ik_p x_3 - i\omega t\right)$ quasi-plane wave traveling in the x_3 –direction

where $P\left(x_1, x_3, \omega\right) = A\exp\left[ik_p x_1 \sin\theta + ik_p x_3\left(1 - \cos\theta\right)\right]$

We can get a more physical interpretation of the mathematical criterion for the paraxial approximation by looking at a plane wave that travels at a small angle from the z-axis. We can write this plane wave in terms of a quasi-plane wave traveling in the z-direction.

Paraxial Approximation



$$P(x_1, x_3, \omega) = A \exp\left[ik_p x_1 \sin\theta + ik_p x_3 (1 - \cos\theta) \right]$$

$$\frac{\partial^2 P}{\partial x_1^2} = -k_p^2 P \sin^2\theta \cong -k_p^2 P \theta^2$$

$$2ik_p \frac{\partial P}{\partial x_3} = -2k_p^2 P (1 - \cos\theta) \cong k_p^2 P \theta^2$$

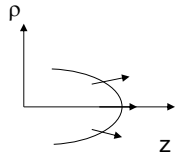
$$\frac{\partial^2 P}{\partial x_3^2} = -k_p^2 P (1 - \cos\theta)^2 \cong -\frac{k_p^2 P \theta^4}{4}$$

if θ <0.5 rad (about 30°)

$$\Rightarrow \quad \left| \frac{\partial^2 P}{\partial x_3^2} \right| << \left| \frac{\partial^2 P}{\partial x_1^2} \right|, \left| 2ik_p \frac{\partial P}{\partial x_3} \right|$$

If we look at the amplitude expression for this quasi-plane wave we see that is the angle is small (typically less than about 30 degrees from the z-axis) then the second derivative of this amplitude with respect to z is much less than the other derivatives shown, agreeing with the condition we stated earlier for the paraxial approximation to be valid. Thus, if a beam of sound is well collimated so that it does not spread too much, the paraxial approximation is valid.

Paraxial Approximation



wave equation (cylindrical coordinates)
with axial symmetry

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial p}{\partial\rho}\right)+\frac{\partial^2 p}{\partial z^2}-\frac{1}{c_p^2}\frac{\partial^2 p}{\partial t^2}=0$$

NOTE: ρ here is a radius, not the density

Let $p=P(\rho,z,\omega)\exp\left(ik_p z-i\omega t\right)$

and assume $\left|\frac{\partial^2 P}{\partial z^2}\right|<<\left|2ik_p\frac{\partial P}{\partial z}\right|,\left|\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial P}{\partial\rho}\right)\right|$

then

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial P}{\partial\rho}\right)+2ik_p\frac{\partial P}{\partial z}=0$$

paraxial wave equation

So far we have been looking at the wave equation in Cartesian or spherical coordinates. If we write the wave equation in cylindrical coordinates and examine a quasi-plane wave traveling in the z-direction, then the form of the paraxial wave equation is as shown.

Paraxial Wave Equation

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial P}{\partial\rho}\right)+2ik_p\frac{\partial P}{\partial z}=0 \qquad p=P(\rho,z,\omega)\exp\left(ik_p z-i\omega t\right)$$

Solutions

$$P=P_0$$ plane wave

$$P=\frac{P_0}{z}\exp\left[ik_p\frac{\rho^2}{2z}\right]$$ spherical wave

$$P=\frac{P_0}{(z-z_0)+iz_c}\exp\left[\frac{ik_p\rho^2}{2R(z)}-\frac{\rho^2}{\left(w(z)\right)^2}\right]$$ Gaussian beam

Here are the exact solutions to this paraxial wave equation for a plane wave, a spherical wave, and a Gaussian beam.

Complex form of the Gaussian beam commonly used

$$p = p_0 \frac{q(0)}{q(z)} \exp\left[ik_p z\right] \exp\left[\frac{ik_p \rho^2}{2q(z)}\right]$$

$$q(z) = q(0) + z \qquad q(0) = -\left(z_0 + \frac{i\pi w_0^2}{\lambda_p}\right)$$

We saw earlier that the Gaussian beam has a rather complex structure. However, we can write it in a much more compact form if we write the Gaussian in terms of a complex q function as shown.

Gaussian beam amplitudes obey the plane wave relationship

$$p = \rho_m c_{pm} v_z$$

↑ density, not radius

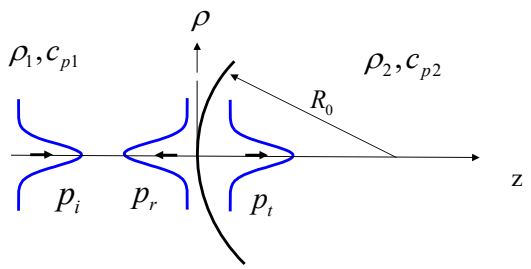
$$p = p_0 \frac{q(0)}{q(z)} \exp\left[ik_{p1} z\right] \exp\left[\frac{ik_{p1}\rho^2}{2q(z)}\right] \qquad k_{p1} = \omega / c_{p1}$$

$$v_z = v_0 \frac{q(0)}{q(z)} \exp\left[ik_{p1} z\right] \exp\left[\frac{ik_{p1}\rho^2}{2q(z)}\right]$$

$$= \frac{p_0}{\rho_1 c_{p1}} \frac{q(0)}{q(z)} \exp\left[ik_{p1} z\right] \exp\left[\frac{ik_{p1}\rho^2}{2q(z)}\right]$$

Note that we can show that the Gaussian beam amplitudes as measured in terms of different quantities such as pressure (or stress) and velocity (or displacement) simply obey the plane wave relationships for those quantities we discussed earlier.

Reflection and Transmission at a Spherical Interface of radius R_0



$$p_i = \tilde{P}_i(z)\exp\left[ik_{p1}z + i\omega t_0\right]\exp\left[\frac{ik_{p1}\rho^2}{2q_i(z)}\right]$$

$$p_r = \tilde{P}_r(z_2)\exp\left[ik_{p1}z_2 + i\omega t_0\right]\exp\left[\frac{ik_{p1}\rho^2}{2q_r(z_2)}\right] \qquad z_2 = -z$$

$$p_t = \tilde{P}_t(z)\exp\left[ik_{p2}z + i\omega t_0\right]\exp\left[\frac{ik_{p2}\rho^2}{2q_t(z)}\right]$$

One of the important properties of Gaussian beams is that we can propagate them through complex geometries and they always remain well-behaved. Consider, for example, the propagation of a Gaussian beam through a spherically curved interface at normal incidence. This is a very specific problem that we can use to show the process. We can write the incident, transmitted, and reflected Gaussians as shown.

Amplitudes are related through plane wave reflection and transmission coefficients

$$\tilde{P}_r(0) = \frac{\rho_2 c_{p2} - \rho_1 c_{p1}}{\rho_1 c_{p1} + \rho_2 c_{p2}} \tilde{P}_i(0) = R_p \tilde{P}_i(0)$$

$$\tilde{P}_2(0) = \frac{2\rho_2 c_{p2}}{\rho_1 c_{p1} + \rho_2 c_{p2}} \tilde{P}_i(0) = T_p \tilde{P}_i(0)$$

Phase terms are related by

transmitted beam

$$\frac{1}{q_t(0)} = \left(\frac{c_{p2}}{c_{p1}} - 1\right)\frac{1}{R_0} + \frac{c_{p2}}{c_{p1}}\frac{1}{q_i(0)}$$

reflected beam

$$\frac{1}{q_r(0)} = \frac{2}{R_0} + \frac{1}{q_i(0)}$$

radius of interface

By matching the amplitudes and phases for pressure and velocity at the interface we can show that the amplitudes at the interface are related through plane wave reflection and transmission coefficients and the phase terms are also related.

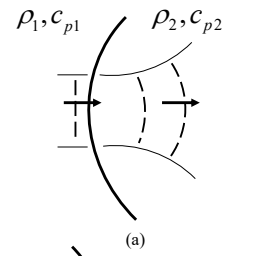
Transmission/reflection laws in terms of beam widths and wave front curvatures

$$w_t(0) = w_r(0) = w_i(0)$$

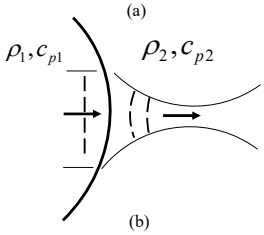
$$\frac{1}{R_t(0)} = \left(\frac{c_{p2}}{c_{p1}} - 1\right)\frac{1}{R_0} + \frac{c_{p2}}{c_{p1}}\frac{1}{R_i(0)}$$

$$\frac{1}{R_r(0)} = \frac{2}{R_0} + \frac{1}{R_i(0)}$$

One finds that the beam widths match at the interface and we can also relate the wave front curvatures.

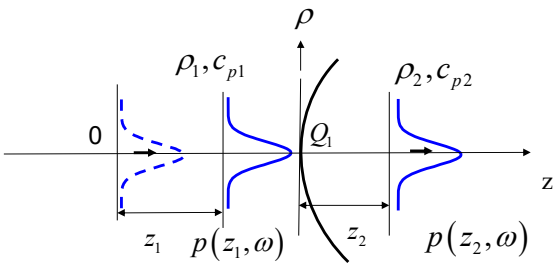


Defocusing interface $c_{p2} > c_{p1}$



Focusing interface $c_{p2} > c_{p1}$

Suppose the beam waist of the incident wave is located at the interface. In this simple case one can easily see the effects of focusing or defocusing depending on the curvature of the interface. Even for a focusing interface the Gaussian beam remains well-behaved in the second medium, unlike a spherical wave which would lead to singularities.



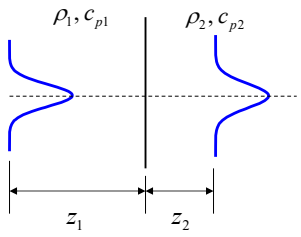
$$q_1(z_1) = q_1(0) + z_1 \qquad p(z_1,\omega) = \tilde{P}(0)\frac{q_1(0)}{q_1(z_1)}\exp(ik_{p1}z_1)\exp\left[\frac{ik_{p1}}{2}\frac{\rho^2}{q_1(z_1)}\right]$$

$$p(z_2,\omega) = \tilde{P}(0)\frac{q_2(Q_1)}{q_2(z_2)}T_{12}\frac{q_1(0)}{q_1(Q_1)}\exp(ik_{p1}z_1 + ik_{p2}z_2)\exp\left[\frac{ik_{p1}}{2}\frac{\rho^2}{q_2(z_2)}\right]$$

$$q_2(z_2) = q_2(Q_1) + z_2 \qquad q_2(Q_1) = \frac{q_1(Q_1)}{\left(\frac{c_{p2}}{c_{p1}} - 1\right)\frac{q_1(Q_1)}{R_0} + \frac{c_{p2}}{c_{p1}}}$$

Consider a Gaussian beam that originates at some distance from the interface. We can write the incident and transmitted Gaussian beams as shown. We could also look at the reflected beam but in NDE problems we are often more concerned with the transmitted wave field.

Gaussian beam transmitted
through a planar interface



$$p(z_2,\omega) = \tilde{P}(0)\frac{q_1(0)}{q_1(0)+\tilde{z}}\exp\left(ik_{p1}z_1 + ik_{p2}z_2\right)\exp\left[\frac{ik_{p1}}{2}\frac{\rho^2}{q_1(0)+\tilde{z}}\right]$$

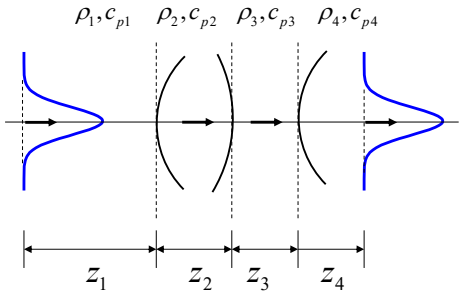
$$\tilde{z} = z_1 + \frac{c_{p2}}{c_{p1}}z_2$$

Gaussian beam in a single medium

$$p = p_0\frac{q(0)}{q_1(0)+z}\exp\left[ik_p z\right]\exp\left[\frac{ik_p}{2}\frac{\rho^2}{q_1(0)+z}\right]$$

Here is the expression for the transmitted Gaussian beam. We see that it has the same form as a Gaussian beam traveling in a single medium where we replace the z distance in the amplitude and Gaussian phase terms by an effective distance, a behavior we have seen before for other problems in the paraxial approximation.

Gaussian beams in multiple media



$$p\left(z_{M+1},\omega\right)=\tilde{P}(0)\frac{q_{M+1}(0)}{q_{M+1}\left(z_{M+1}\right)}\prod_{m=1}^{M}\tilde{T}_{m\,m+1}\frac{q_m(0)}{q_m\left(z_m\right)}$$
$$\cdot\exp\left[i\sum_{m=1}^{M+1}k_{pm}z_m\right]\exp\left[\frac{ik_{pM+1}}{2}\frac{\rho^2}{q_{M+1}\left(z_{M+1}\right)}\right]$$

In fact, we can easily write down the wave field of the Gaussian after propagation through multiple media. Multiple transmission coefficients are present but the form of the response remains very similar.

Propagation Laws $q_m(z_m) = q_m(0) + z_m$

Transmission Laws $$q_{m+1}(Q_m) = \frac{q_m(Q_m)}{\left(\frac{c_{p\,m+1}}{c_{p\,m}} - 1\right)\frac{q_m(Q_m)}{(R_0)_m} + \frac{c_{p\,m+1}}{c_{p\,m}}}$$

Reflection Laws $$q_{m+1}(Q_m) = \frac{q_m(Q_m)}{\frac{2}{(R_0)_m} q_m(Q_m) + 1}$$

We have seen that the q(z) function plays a key role in Gaussian beam theory. We can write down explicit propagation, transmission, and reflection laws in terms of this function.

ABCD Matrix Forms of the Laws

$$q_f = \frac{Aq_i + B}{Cq_i + D}$$

Propagation Laws

$$\begin{bmatrix} A^d & B^d \\ C^d & D^d \end{bmatrix} = \begin{bmatrix} 1 & z_m \\ 0 & 1 \end{bmatrix}$$

Transmission Laws

$$\begin{bmatrix} A^t & B^t \\ C^t & D^t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{\left(c_{p\,m+1}/c_{p\,m} - 1\right)}{\left(R_0\right)_m} & \frac{c_{p\,m+1}}{c_{p\,m}} \end{bmatrix}$$

Reflection Laws

$$\begin{bmatrix} A^r & B^r \\ C^r & D^r \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{2}{\left(R_0\right)_m} & 1 \end{bmatrix}$$

In fact, we can write all these laws in a common form in terms of the elements of a 2x2 ABCD matrix.

Gaussian beams in multiple media

$$p(z_{M+1},\omega) = \mathcal{T}\tilde{P}(0)\frac{q_1(0)}{A^G q_1(0)+B^G}\exp\left[i\omega t_0\right]\exp\left[\frac{ik_{pM+1}}{2}\frac{\rho^2}{\frac{A^G q_1(0)+B^G}{C^G q_1(0)+D^G}}\right]$$

$$\mathcal{T} = \prod_{m=1}^{M}\tilde{T}_{m\,m+1} \qquad t_0 = \sum_{m=1}^{M+1} z_m / c_{pm}$$

transmission or reflection coefficients

$$\begin{bmatrix} A^G & B^G \\ C^G & D^G \end{bmatrix} = \begin{bmatrix} A^d_{M+1} & B^d_{M+1} \\ C^d_{M+1} & D^d_{M+1} \end{bmatrix}\begin{bmatrix} A^t_M & B^t_M \\ C^t_M & D^t_M \end{bmatrix}\cdots\begin{bmatrix} A^t_1 & B^t_1 \\ C^t_1 & D^t_1 \end{bmatrix}\begin{bmatrix} A^d_1 & B^d_1 \\ C^d_1 & D^d_1 \end{bmatrix}$$

After transmission or reflection from multiple interfaces the Gaussian beam retains a simple form in terms of a single ABCD matrix that is obtained simply by multiplying all the individual ABCD matrices together.

Gaussian beam after propagating through M+1 media

$$p(z_{M+1},\omega)=\mathcal{F}\tilde{P}(0)\frac{q_1(0)}{A^G q_1(0)+B^G}\exp\left[i\omega t_0\right]\exp\left[\frac{ik_{pM+1}}{2}\frac{\rho^2}{\frac{A^G q_1(0)+B^G}{C^G q_1(0)+D^G}}\right]$$

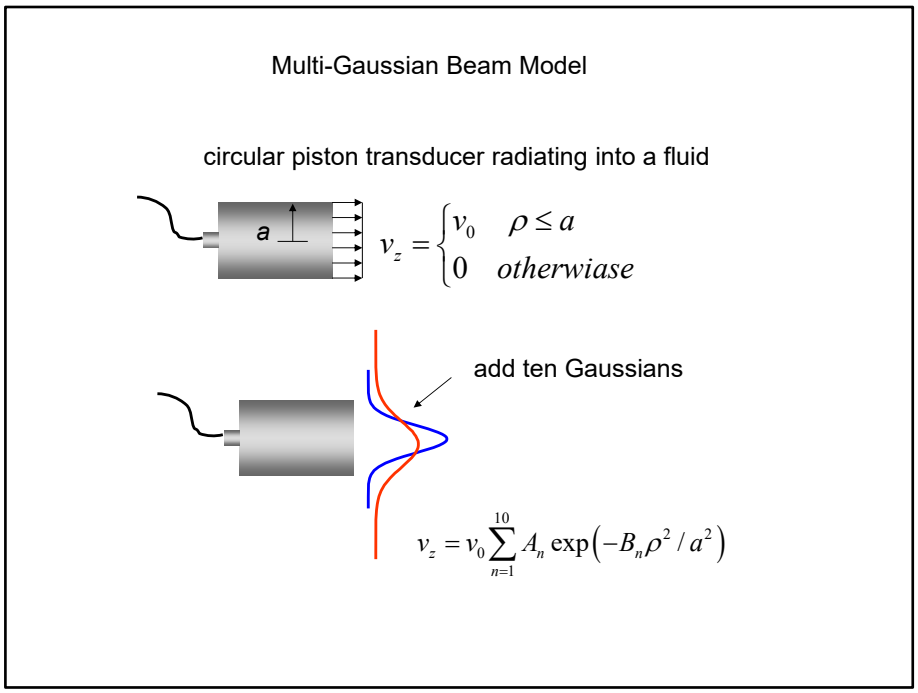
Gaussian beam in single medium

$$p(z_1,\omega)=\tilde{P}(0)\frac{q_1(0)}{A_1^d q_1(0)+B_1^d}\exp\left[ik_{p1}z_1\right]\exp\left[\frac{ik_{p1}}{2}\frac{\rho^2}{\frac{A_1^d q_1(0)+B_1^d}{C_1^d q_1(0)+D_1^d}}\right]$$

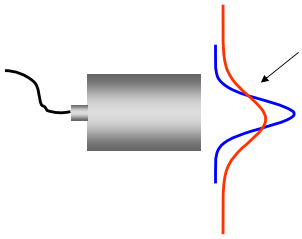
$$=\tilde{P}(0)\frac{q_1(0)}{q_1(0)+z_1}\exp\left[ik_{p1}z_1\right]\exp\left[\frac{ik_{p1}}{2}\frac{\rho^2}{q_1(0)+z_1}\right]$$

$$\begin{bmatrix}A_1^d & B_1^d\\ C_1^d & D_1^d\end{bmatrix}=\begin{bmatrix}1 & z_1\\ 0 & 1\end{bmatrix}$$

Here we see how the Gaussian beam retains its general form even in complex problems in terms of the ABCD matrices. This is for the very special case of normal incidence on the multiple interfaces present but a similar behavior is present in more general cases.



Gaussian beams can be used to develop a very efficient beam model for a circular piston transducer. Wen and Breazeale showed that one can simply superimpose as few as 10 different Gaussians on the face of the transducer to produce the same wave field that is generated by a uniform velocity (piston) transducer. These Gaussians will then produce ten propagating Gaussian beams that we can be propagated in the manner we have just discussed. This is a **multi-Gaussian beam model**.



$$p = \sum_{n=1}^{10} \frac{\tilde{P}_n(0)}{\left[q_1(0)\right]_n + z_1} \exp\left[ik_p z\right] \exp\left[\frac{ik_p \rho^2}{2} \frac{1}{\left[q_1(0)\right]_n + z_1}\right]$$

$$\tilde{P}_n(0) = \rho c_p v_0 A_n$$

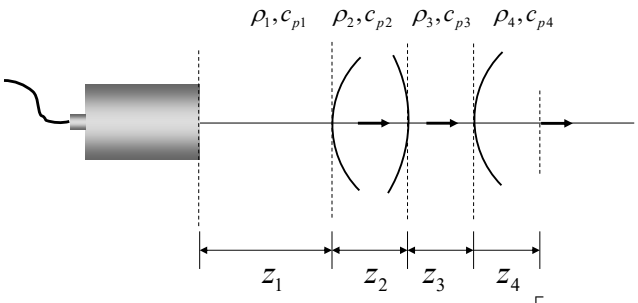
$$\left[q_1(0)\right]_n = \frac{-ik_p a^2/2}{B_n}$$

Here, for example, is a multi-Gaussian beam model for a single medium. It is expressed in terms of ten known complex-valued coefficients A_n , B_n .

```
function [a, b] = gauss_c
%  gauss_c returns the coefficients developed by Wen and Breazeale to model a
%  piston transducer
a = zeros(10,1);
b = zeros(10,1);
a(1) = 11.428 + 0.95175*i;
a(2) = 0.06002 - 0.08013*i;
a(3) = -4.2743 - 8.5562*i;
a(4) = 1.6576 +2.7015*i;
a(5) = -5.0418 + 3.2488*i;
a(6) = 1.1227 - 0.68854*i;
a(7) = -1.0106 - 0.26955*i;
a(8) = -2.5974 + 3.2202*i;
a(9) = -0.14840 -0.31193*i;
a(10) = -0.20850 - 0.23851*i;
b(1) = 4.0697 + 0.22726*i;
b(2) = 1.1531 - 20.933*i;
b(3) = 4.4608 + 5.1268*i;
b(4) = 4.3521 +14.997*i;
b(5) = 4.5443 + 10.003*i;
b(6) = 3.8478 + 20.078*i;
b(7) = 2.5280 -10.310*i;
b(8) = 3.3197 - 4.8008*i;
b(9) = 1.9002 - 15.820*i;
b(10) = 2.6340 + 25.009*i;
```

Here are the ten coefficients obtained by Wen and Breazeale.

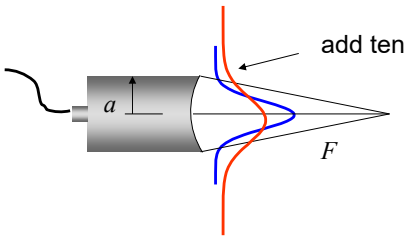
Multi-Gaussian beam model of a piston transducer in multiple media



$$p\left(z_{M+1},\omega\right)=\sum_{n=1}^{10}\mathcal{T}\tilde{P}_n\left(0\right)\frac{\left[q_1\left(0\right)\right]_n}{A^G\left[q_1\left(0\right)\right]_n+B^G}\exp\left[i\omega t_0\right]\exp\left[\frac{ik_{p\,M+1}}{2}\frac{\rho^2}{\frac{A^G\left[q_1\left(0\right)\right]_n+B^G}{C^G\left[q_1\left(0\right)\right]_n+D^G}}\right]$$

These ten Gaussians can then be transmitted or reflected multiple times to obtain the Gaussian beam in the final medium.

multi-Gaussian beam model for a
spherically focused circular transducer



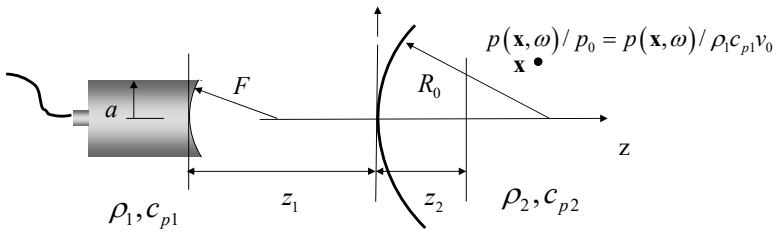
$$v_z = v_0 \sum_{n=1}^{10} A_n \exp\left(-B_n \rho^2 / a^2\right) \exp\left(-ik\rho^2 / 2F\right)$$

$$= v_0 \sum_{n=1}^{10} A_n \exp\left(-\tilde{B}_n \rho^2 / a^2\right)$$

$$\tilde{B}_n = B_n + \frac{ika^2}{2F}$$

A multi-Gaussian bean can also easily model a spherically focused transducer by simply including an additional phase term at the transducer face, as previously discussed. This term can be easily included by simply changing the B_n coefficients in the multi-Gaussian model.

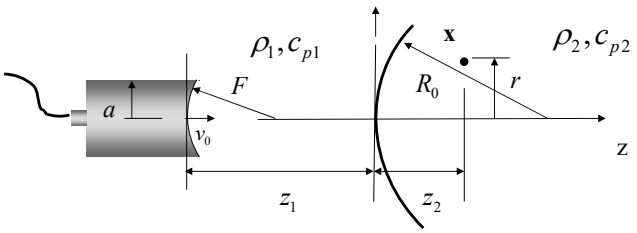
MATLAB multi- Gaussian Beam Model



The time delay factor $\exp\left(ik_{p1}z_1 + ik_{p2}z_2\right)$ is not included in this model

```
% MGB_script
% This script calculates the normalized
% pressure, p/p0, due to a planar piston transducer radiating at normal
% incidence through a spherically curved interface.
% This is a multi-Gaussian beam model that uses the 10 coefficients
%  of Wen and Breazeale to calculate the wave field.

clear    % clear the workspace so that any previous parameter choices
% are eliminated
```

We will illustrate a multi-Gaussian beam model by considering the radiation of a planar or spherically focused piston transducer through a spherically curved interface at normal incidence, as shown above. A MATLAB script, MGB_script, will be developed explicitly for this problem. The script first clears the workspace so that any parameters remaining from a previous calculation are removed.



$q_1(0)$ $q_1(z_1)$ $q_2(0)$ $q_2(z_2)$

$$q_1(z_1) = q_1(0) + z_1 \qquad q_2(0) = \frac{q_1(z_1)}{\left(\frac{c_{p2}}{c_{p1}} - 1\right)\frac{q_1(z_1)}{R_0} + \frac{c_{p2}}{c_{p1}}} \qquad q_2(z_2) = q_2(0) + z_2$$

$$\frac{p(\mathbf{x},\omega)}{\rho_1 c_{p1} v_0} = P(0) T_{12} \frac{q_1(0)}{q_1(z_1)} \frac{q_2(0)}{q_2(z_2)} \underbrace{\exp\left[ik_{p1}z_1 + ik_{p2}z_2\right]}_{\text{omit}} \exp\left[\frac{ik_{p2}}{2}\frac{r^2}{q_2(z_2)}\right]$$

$$P(0) = A_n \qquad q_1(0) = \frac{-ik_{p1}a^2}{2B_n}$$

Here is the explicit expression we will use and the definition of all the underlying parameters. We do not use ABCD matrices here since it is a very simple case. We will not include the propagation phase term here since that term simply provides a time delay term for the arriving waves but does not affect their spatial distribution. The q_1 function describes the behavior of the Gaussian beam in the first medium and the q_2 function describes its behavior in the second medium. T_{12} is the plane wave transmission based on pressure ratios. Shown is a single term in a ten-term multi-Gaussian beam model.

MATLAB multi- Gaussian Beam Model

```
% get input parameters
f = 5;                          %frequency  (MHz)
a = 6.35;                       % transducer radius (mm)
Fl = inf;                       % transducer focal length (mm)
z1 = 0;                         % path length in medium 1 (mm)
z2 = linspace(6, 125, 200);         % path length in medium 2 (mm)
r =0.0;                         % distance from ray axis (mm)
R0 = inf;                       % interface radius of curvature (mm)
d1 = 1.0;                       % density, medium 1 (gm/cm^3)
d2 =1.0;                        % density, medium 2 (gm/cm^3)
c1 = 1480.;                     % wave speed, medium 1 (m/sec)
c2 = 1480.;                     % wave speed, medium 2 (m/sec)
```

These parameters are for a ½ inch diameter, 5 MHz planar piston transducer radiating into water (no interface). The on-axis pressure values from 6 to 125 mm are being sought.

Here are the input parameters in the script. The default case is where we are radiating into a single medium and calculating the on-axis normalized pressure from 6 mm to 125 mm on the transducer axis for a planar piston transducer.

MATLAB multi- Gaussian Beam Model

```
% get Wen and Breazeale coefficients (10)
[A, B] = gauss_c;

% transmission coefficient (based on pressure ratio)

T = (2*c2*d2)/(c1*d1+c2*d2);

h = 1/R0;                                   % interface curvature

zr = eps*(f == 0) + 1000*pi*(a^2)*f./c1;    % "Rayleigh" distance, ka^2/2
k1 = 2*pi*1000*f./c1;                       % wave number in medium 1
```

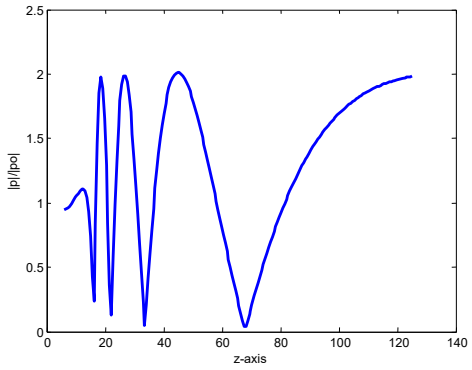
We also need the Wen and Breazeale coefficients and we define curvature of the interface in terms of the radius. Two parameters, the Rayleigh distance and the wave number associated with the first medium are also defined.

MATLAB multi- Gaussian Beam Model

```
%initialize  pressure to zero
p = 0;

%multi-Gaussian beam model
for  j = 1:10      % form up multi-Gaussian beam model with
                   % 10 Wen and Breazeale coefficients

 b =B(j) + i*zr./Fl;                    % modify coefficients for focused probe

q = z1 - i*zr./b;
K = q.*(1 -(c1/c2));
M = (1 +K.*h);
ZR = q./M;
m = 1./(ZR +(c2/c1).*z2);
   t1 = A(j)./(1 + (i.*b./zr).*z1);
   t2 = t1.*T.*ZR.*m;
   p = p + t2.*exp(i.*(k1./2).*m.*(r.^2));

end
```

We then simply add up ten Gaussian beams of the form shown previously, changing the Wen and Breazeale coefficients for the focused case (Fl = inf for a planar interface) to compute the normalized pressure.

```
>> MGB_script

%plot magnitude of on-axis pressure
plot(z2, abs(p))
xlabel('z-axis')
ylabel('|p|/|po|')
```



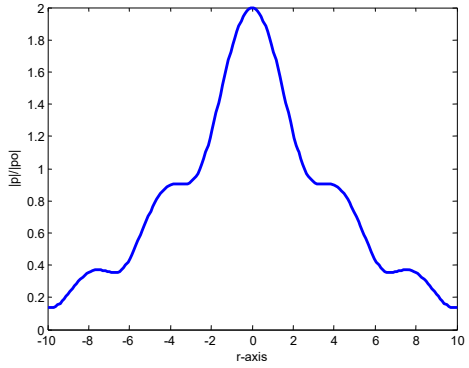
After executing the MGB_script, we plot the on axis pressure, showing the near field behavior of the planar piston transducer.

Cross-axis plot at z = 130 mm (near last maximum)

```
% get input parameters
f = 5;                          %frequency  (MHz)
a = 6.35;                       % transducer radius (mm)
FI = inf;                       % transducer focal length (mm)
z1 = 0;                         % path length in medium 1(mm)
z2 = 130;                       %  path length in medium 2 (mm)
r =linspace(-10, 10, 200);      % distance from ray axis (mm)
R0 = inf;                       % interface radius of curvature (mm)
d1 = 1.0;                       % density, medium 1 (gm/cm^3)
d2 =1.0;                        % density, medium 2 (gm/cm^3)
c1 = 1480.;                     % wave speed, medium 1 (m/sec)
c2 = 1480.;                     % wave speed, medium 2 (m/sec)
```

Now, let us change the input parameters in the script (changes in red) and plot the pressure versus radius at a distance near the last on-axis maximum. Save the script with these changes and execute.

```
>> MGB_script

%plot magnitude of on-axis pressure
plot(r, abs(p))
xlabel('r-axis')
ylabel('|p|/|po|')
```



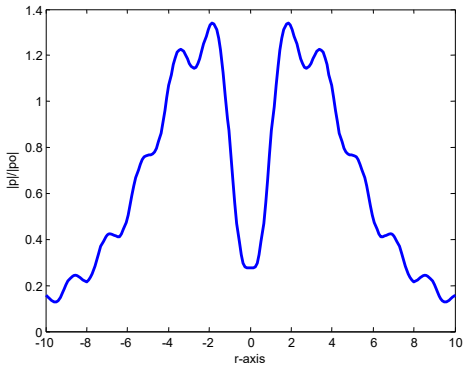
Here is a plot of the off-axis pressure versus the radius near that last maximum in the near field, showing a main lobe and some side lobes for the beam

Cross-axis plot at z = 65 mm (near last on-axis null)

```
% get input parameters
f = 5;                          %frequency  (MHz)
a = 6.35;                       % transducer radius (mm)
FI = inf;                       % transducer focal length (mm)
z1 = 0;                         % path length in medium 1 (mm)
z2 = 65;                        % path length in medium 2  (mm)
r =linspace(-10, 10, 200);      % distance from ray axis (mm)
R0 = inf;                       % interface radius of curvature (mm)
d1 = 1.0;                       % density, medium 1 (gm/cm^3)
d2 =1.0;                        % density, medium 2 (gm/cm^3)
c1 = 1480.;                     % wave speed, medium 1 (m/sec)
c2 = 1480.;                     % wave speed, medium 2 (m/sec)
```

Now, let us look instead at the cross axis behavior at a distance close to the last on-axis null in the near field.

```
>> MGB_script

%plot magnitude of on-axis pressure
plot(r, abs(p))
xlabel('r-axis')
ylabel('|p|/|po|')
```



We see pressure is small near the null on the axis but has large values off axis.

Now, generate an image of the entire beam

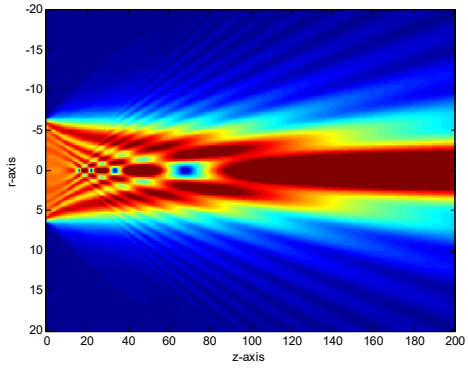
```
% get input parameters
f = 5;                          %frequency  (MHz)
a = 6.35;                       % transducer radius (mm)
Fl = inf;                % transducer focal length (mm)
z1 = 0;                     % path length in medium 1 (mm)
z2t = linspace(0,200,500);
rt =linspace(-20, 20, 200);
[z2, r]=meshgrid(z2t, rt);
R0 = inf;                       % interface radius of curvature (mm)
d1 = 1.0;                       % density, medium 1 (gm/cm^3)
d2 =1.0;                        % density, medium 2 (gm/cm^3)
c1 = 1480.;                     % wave speed, medium 1 (m/sec)
c2 = 1480.;                     % wave speed, medium 2 (m/sec)
```

To see a cross section of the entire beam we need to change the script to generate a grid of points in the two-dimensional (r, z) plane.

>> MGB_script

```
%plot magnitude of the pressure
image(z2t, rt, abs(p)*50)
xlabel('z-axis')
ylabel('r-axis')
```

Note: unequal scales



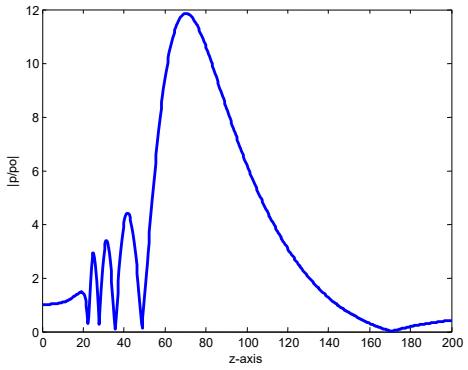
Here is an 2-D image of the transducer wave field showing its complex behavior (at a given frequency) in the near field. The magnitude of the pressure has been multiplied by a factor here so the colors in the image give a good visual of the wave field.

We can also examine a 10 MHz, 76.2 mm focal length focused transducer

```
% get input parameters
f = 10;                          %frequency  (MHz)
a = 6.35;                        % transducer radius (mm)
Fl = 76.2;                       % transducer focal length (mm)
z1 = 0;                          % path length in medium 1(mm)
z2 = linspace(0,200,500);        % path length in medium 2 (mm)
r=0;                             % distance from ray axis (mm)
R0 = inf;                        % interface radius of curvature (mm)
d1 = 1.0;                        % density, medium 1 (gm/cm^3)
d2 =1.0;                         % density, medium 2 (gm/cm^3)
c1 = 1480.;                      % wave speed, medium 1 (m/sec)
c2 = 1480.;                      % wave speed, medium 2 (m/sec)
```

Now let us consider a 10 MHz spherically focused transducer radiating into water and examine the on-axis behavior.

```
>> MGB_script

%plot magnitude of on-axis pressure
plot(z2, abs(p))
xlabel('z-axis')
ylabel('|p/po|')
```



Here is the on-axis plot, showing the large pressure near the focus, the near-field behavior, and the existence of the null (shown previously) at a distance greater than the focal length.

Now, plot an image of this focused transducer wave field

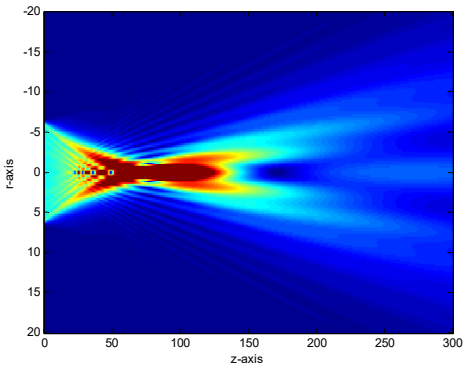
```
% get input parameters
f = 10;                    %frequency  (MHz)
a = 6.35;                  % transducer radius (mm)
Fl = 76.2;                 % transducer focal length (mm)
z1 = 0;                    % path length in medium 1 (mm)
z2t = linspace(0,300,500);
rt=linspace(-20,20,200);
[z2, r]=meshgrid(z2t, rt);
R0 = inf;                        % interface radius of curvature (mm)
d1 = 1.0;                        % density, medium 1 (gm/cm^3)
d2 =1.0;                   % density, medium 2 (gm/cm^3)
c1 = 1480.;                % wave speed, medium 1 (m/sec)
c2 = 1480.;                % wave speed, medium 2 (m/sec)
```

Here we change the script to show a 2-D cross-sectional image.

>> MGB_script

```
%plot magnitude of the pressure
image(z2t, rt, abs(p)*25)
xlabel('z-axis')
ylabel('r-axis')
```

Note: unequal scales



Comparing with the previous planar case we clearly see the effects of focusing. We can consider many other cases but we will stop here.

Homework problems

F.2, F.3

Homework problems from Appendix F.

Flaw Scattering Models

Having examined the waves generated by transducers we now need to consider the scattered waves generated when an incident wave strikes a flaw.

Learning Objectives

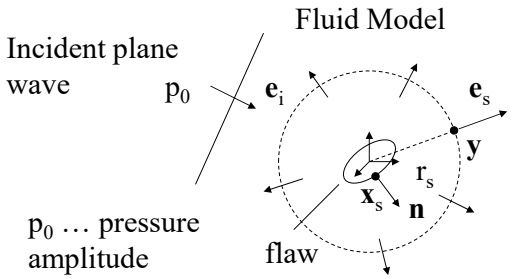
Far-field scattering amplitude

Kirchhoff approximation
Born approximation
Separation of Variables

Examples of scattering of simple shapes (spherical pore, flat crack, side-drilled hole)

We will see that the waves scattered from a flaw can be described in terms of a far field scattering amplitude and examine three different methods for modeling those scattered waves. We will examine scattering for some of the simple shapes commonly used as reference “flaws”.

Flaw Scattering



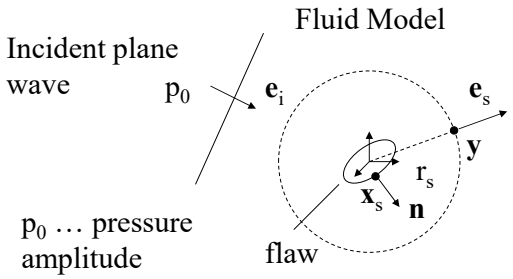
At many wavelengths away from the flaw the scattered waves are spherical waves

$$p^{scatt}(\mathbf{y},\omega) = p_0 A(\mathbf{e}_i;\mathbf{e}_s)\frac{\exp(ikr_s)}{r_s}$$ scattered pressure

A is called the plane wave far-field scattering amplitude

In most NDE inspections the flaw being examined will be located at many wavelengths from the receiving transducer. In that case the flaw acts like a point source having a complicated radiation pattern, i.e. it radiates a spherical wave with an amplitude coefficient that is called the far field scattering amplitude. Shown is the case for a scatterer in a fluid where the incident wave is a plane wave and A is the **plane wave far field scattering amplitude**. A is a function of both the incident and scattering directions of the waves as well as the frequency but we will typically show the unit vectors $\mathbf{e}_i$ and $\mathbf{e}_s$ that define the incident and scattered directions in the arguments of A but not the frequency. Note that A has the dimensions of a length.

Flaw Scattering



plane wave far field scattering amplitude of the flaw

$$A(\mathbf{e}_i;\mathbf{e}_s)=\frac{-1}{4\pi}\int_{S_f}\left[\frac{\partial\tilde{p}}{\partial n}+ik(\mathbf{e}_s\cdot\mathbf{n})\tilde{p}\right]\exp(-ik\mathbf{x}_s\cdot\mathbf{e}_s)\,dS(\mathbf{x}_s)$$

$$\frac{\partial p}{\partial n}=-i\omega\rho v_n \qquad \tilde{p}=\frac{p(\mathbf{x}_s,\omega)}{p_0}$$

In the fluid case just described we can actually write down an integral expression for the plane wave far field scattering amplitude in terms of the normalized pressure and its normal derivative (the normal derivative of the pressure is just proportional to the normal velocity, as shown), where the integral is taken over the surface of the flaw. The unit vector $\mathbf{e}_s$ here is a unit vector in the scattering direction being examined.

Flaw Scattering

$$A(\mathbf{e}_i;\mathbf{e}_s)=\frac{-1}{4\pi}\int_{S_f}\left[\frac{\partial\tilde{p}}{\partial n}+ik(\mathbf{e}_s\cdot\mathbf{n})\tilde{p}\right]\exp(-ik\mathbf{x}_s\cdot\mathbf{e}_s)\,dS(\mathbf{x}_s)$$

$$\tilde{p}=\frac{p(\mathbf{x}_s,\omega)}{p_0}$$

To obtain the far field scattering amplitude, we need to know the pressure and velocity on the surface of the flaw due to the incident and scattered waves

These quantities can be found by solving a flaw scattering boundary value problem.

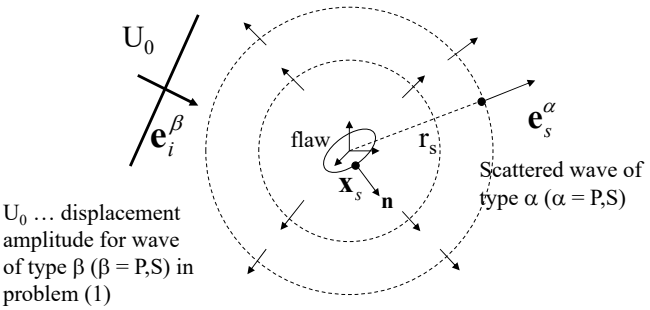
To describe the scattering amplitude we need both the pressure and its normal derivative (or, equivalently, the normal velocity) on the flaw surface which can be found by solving a boundary value problem. To avoid having to go into that detail we can try to use approximations to obtain these surface fields. Shortly, we will examine such approximations

Flaw Scattering

Elastic Solid

$$\mathbf{u}^{scatt}(\mathbf{y},\omega) = U_0 \frac{\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{P}\right)}{r_s}\exp\left(ik_p r_s\right) + U_0 \frac{\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{S}\right)}{r_s}\exp\left(ik_s r_s\right)$$

scattered displacement from a flaw — P-wave — S-wave



U_0 … displacement amplitude for wave of type β (β = P,S) in problem (1)

If we look at flaw scattering in an elastic solid we see much the same picture as in the fluid case but now at many wavelengths the scattered waves are both spherical P-waves and S-waves and the scattering amplitudes are vectors since they must represent a vector field such as the scattered displacements. The incident waves themselves also can be either P-waves or S-waves.

Flaw Scattering

$\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{\alpha}\right)$ Vector scattering amplitude for a scattered wave of type α due to a plane wave of unit displacement amplitude and type β

$$\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{P}\right)=\left(\mathbf{f}^{P;\beta}\cdot\mathbf{e}_s^{P}\right)\mathbf{e}_s^{P} \qquad (\beta=P,S)$$

$$\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{S}\right)=\left[\mathbf{f}^{S;\beta}-\left(\mathbf{f}^{S;\beta}\cdot\mathbf{e}_s^{S}\right)\mathbf{e}_s^{S}\right]$$

elastic constants

$$\mathbf{f}^{\alpha;\beta}=f_l^{\alpha;\beta}\mathbf{e}_l=-\frac{C_{lkpj}\mathbf{e}_l}{4\pi\rho c_\alpha^2}\int_S\left[\left(\frac{\partial\tilde{u}_p}{\partial x_j}\right)n_k+ik_\alpha e_{sk}^{\alpha}n_p\tilde{u}_j\right]\exp\left(-ik_\alpha\mathbf{x}_s\cdot\mathbf{e}_s^{\alpha}\right)dS\left(\mathbf{x}_s\right)$$

displacements and displacement gradients

$(\alpha=P,S)$ $(\beta=P,S)$

We can express both the scattered P- and S-waves in terms of a vector function **f** which can be represented as an integral over the surface of the flaw of both the displacements and the displacement gradients. These surface fields again can be found by solving a particular boundary value problem or through approximations.

Flaw Scattering

3-D scattering amplitude

$$\mathbf{u}^{scatt}(\mathbf{y},\omega) = U_0 \frac{\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{P}\right)}{r_s}\exp\left(ik_p r_s\right) + U_0 \frac{\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{S}\right)}{r_s}\exp\left(ik_s r_s\right)$$

The received voltage in some measurement models is related to a specific component of the vector scattering amplitude given by

$$A(\omega) = A\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{\alpha}\right) = \mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{\alpha}\right)\cdot\left(-\mathbf{d}^{\alpha}\right)$$

polarization vector of an incident wave from receiving transducer (to be discussed shortly)

The received voltage in a measurement is a scalar quantity so that if it is related to the scattering amplitude, it must involve a specific component of the vector scattering amplitude, which we call A. We will see shortly that this scalar component is obtained by taking the dot product of the vector scattering amplitude with a polarization vector associated with the receiving transducer when it acts as a transmitter. We will describe this component later.

Flaw Scattering

$A(\omega)$ can be obtained from:

Numerical methods
- Separation of variables
- Boundary Elements (BEM)
- Finite Elements (FEM)
- Method of Optimal Truncation (MOOT)
- T-matrix

Approximate Methods
- Kirchhoff Approximation
- Born Approximation

or $A(\omega)$ can be found experimentally through deconvolution: $V_R(\omega) = E(\omega)A(\omega) \rightarrow A(\omega) = \frac{V_R(\omega)E^*(\omega)}{|E(\omega)|^2 + n^2}$

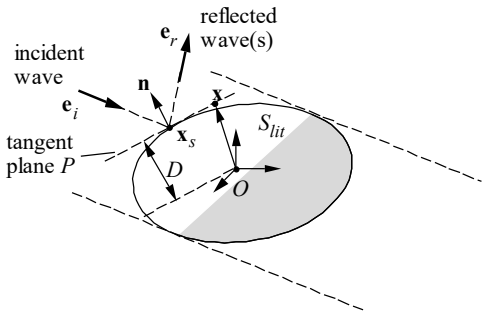
measure model, measure

If we solve a boundary value problem whose solution gives the surface fields needed to define the scattering amplitude, that solution is normally obtained with numerical methods, a number of which are shown above. Here, we will discuss only one of those methods, the method of separation of variables. Alternatively, we can use approximate methods. We will talk about two of those approximations – the Kirchhoff approximation and the Born Approximation.

There is also an experimental way to obtain the scattering amplitude. If we measure the voltage response of the flaw in a setup where we can model or measure all of the other parts of the system, then we can obtain the scattering amplitude through deconvolution, which is normally done, as shown above, with a Wiener filter.

Flaw Scattering

Kirchhoff Approximation for a Volumetric Flaw



on the "lit" surface: fields are those from plane waves interacting with a plane surface

on the rest of the surface the fields are assumed to be zero

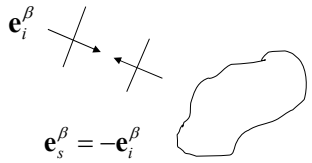
First, consider the Kirchhoff approximation. This is an approximation suitable for high frequencies where the fields on the portion of the surface where the waves directly strike the flaw (called the “lit” surface) are obtained by calculating the interaction of the incident wave (taken to be a plane wave) with a planar surface whose unit normal coincides with the normal to the surface. This plane wave interaction problem can be solved exactly so we obtain the surface fields over the entire lit surface. Over the remaining of the surface the fields are assumed to be identically zero.

Flaw Scattering

It has been recently shown that the pulse-echo scattering amplitude response, $A(\omega)$, for a stress-free flaw (void or crack) in an elastic solid is identical to the scalar (fluid) scattering amplitude:

$$A(\omega) = A\left(\mathbf{e}_i^\beta; -\mathbf{e}_i^\beta\right) = \frac{-ik_\beta}{2\pi} \int_{S_{lit}} \left(\mathbf{e}_i^\beta \cdot \mathbf{n}\right) \exp\left(2ik_\beta \mathbf{x}_s \cdot \mathbf{e}_i^\beta\right) dS\left(\mathbf{x}_s\right)$$

$$\beta = P, S$$

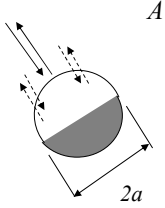


Although in general the scattered response in an elastic solid is quite different from that of a flaw in a fluid, it has been shown that for a stress-free flaw (crack or void) the pulse-echo response for the scattering amplitude is the same for both the fluid and elastic problems yielding an explicit integral over the lit surface that can be evaluated. This allows us to use the simple fluid model even in the elastic wave case.

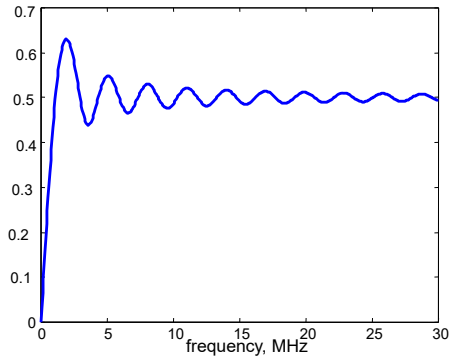
Flaw Scattering-Spherical Void

For the pulse-echo response of a spherical void of radius *a,* fluid model

$$A(\omega) \equiv A(\mathbf{e}_i; -\mathbf{e}_i, \omega) = \frac{-a}{2}\exp(-ika)\left[\exp(-ika) - \frac{\sin(ka)}{ka}\right]$$



|A(ω)|

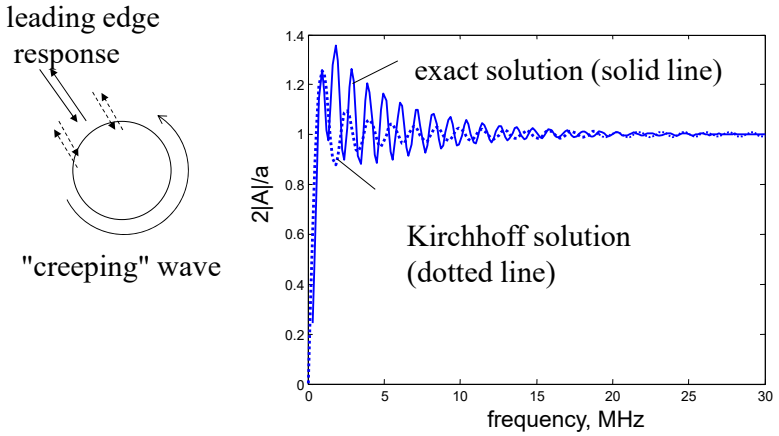


a = 1 mm
c = 5900m/s

Performing the scattering amplitude integral exactly for a spherical void, we can plot the magnitude of the pulse-echo scattering amplitude response as a function of the frequency.

Flaw Scattering-Spherical Void

Comparison with Kirchhoff solution with
exact separation of variables solution for the P-P scattering
amplitude of a void in an elastic solid (pulse-echo)



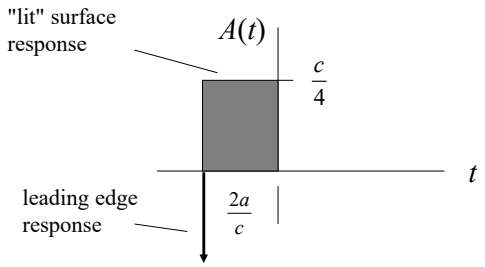
We can also calculate the P-P pulse echo scattering amplitude of the spherical void in an elastic solid numerically (i.e. without approximation) and compare it with the Kirchhoff approximation. We see the two responses have many similarities in terms of their general amplitudes but the oscillations in the responses are different. This difference arises from the fact that the oscillations in the Kirchhoff approximation come from the interference of a large reflection from a point on the surface where the wave first strikes the void (called the **leading edge response**) with the reflections from the remainder of the lit surface. In contrast, the exact separation of variables solution oscillations come from an interference of the leading edge response with a so-called **creeping wave** that travels around the flaw and returns in a direction opposite to the incident wave direction in the pulse-echo setup. The leading edge response is contained in both the separation of variables and Kirchhoff approximation solutions but the creeping wave is not predicted by the Kirchhoff approximation.

Flaw Scattering-Spherical Void

Inverting the scattering amplitude (fluid model) predicted by the Kirchhoff approximation into the time domain

$$A(t) \equiv A(\mathbf{e}_i; -\mathbf{e}_i, t) = \frac{-a}{2}\left[\delta\left(t + \frac{2a}{c}\right) - \frac{c}{2a} U\left(\frac{-2a}{c}, 0; t\right)\right]$$

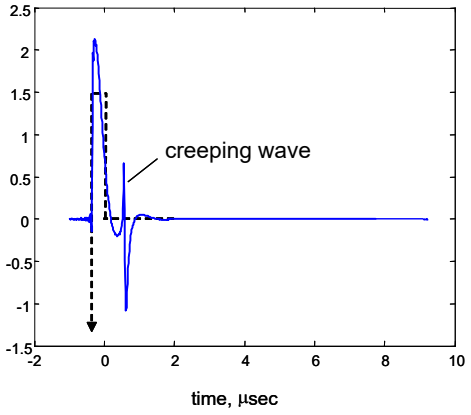
$$U(t_1, t_2; t) = \begin{cases} 1 & t_1 < t < t_2 \\ 0 & otherwise \end{cases}$$



If we invert the Kirchhoff approximation solution into the time domain with an inverse Fourier transform, we see a delta function leading edge response followed by a box-like function over the lit surface (i.e. the front half of the sphere).

Flaw Scattering-Spherical Void

Exact solution
for a void in an
elastic solid (time
domain)
vs the Kirchhoff
approximation solution

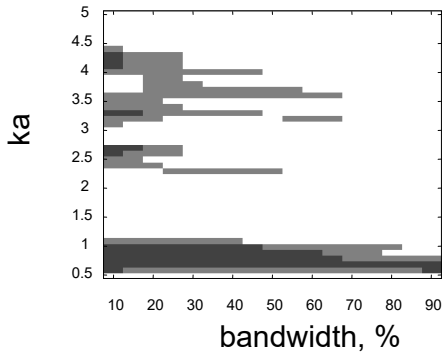


In contrast, we see here the separation of variables solution (plotted in blue) where the leading edge delta function has been removed since it is identical with the delta function obtained from the Kirchhoff approximation. By removing the delta function the remaining waves frequency content is small at high frequencies so we easily perform an inverse FFT to obtain the time-domain waveforms. In addition to the response over the lit surface (which is not a constant) we see a later arriving wave which is the creeping wave. The Kirchhoff approximation is shown as the dotted line response, which includes (symbolically) the delta function .

Flaw Scattering-Spherical Void

Although the Kirchhoff approximation is formally a high frequency approximation (ka >>1), numerical tests have shown it capable of producing good agreement (<1 dB difference) for the peak-to-peak time domain response of the spherical void down to ka =1 provided the bandwidth is sufficient

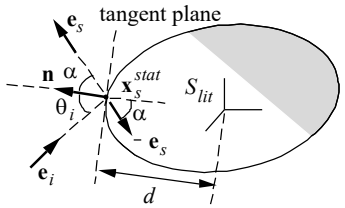
white: differences <1 dB
gray: >1 dB but < 1.5 dB
black: > 1.5 dB



Comparisons of the Kirchhoff approximation with the separation of variables solution show that the Kirchhoff approximation, while it is formally a high frequency approximation (ka >>1) ,where k is the wave number and a is the flaw radius, actually models the amplitude of the pulse-echo signal quite well down to ka =1 provided the bandwidth is sufficiently large

Flaw Scattering-Inclusion

Leading Edge Response - General Convex Inclusion
(scattered mode same as incident mode)



High frequency, stationary phase approximation

$$\Rightarrow \quad \theta_i = \alpha$$

$$A(\mathbf{e}_i;\mathbf{e}_s,\omega) = R_{12}\frac{\sqrt{R_1 R_2}}{2}\exp\left[-ik\left|\mathbf{e}_i - \mathbf{e}_s\right| d\right]$$

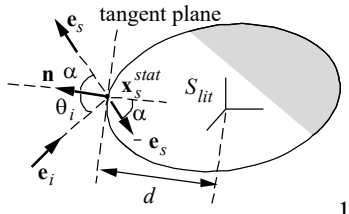
Plane wave reflection coefficient

Gaussian curvature of flaw surface at stationary phase point

The **leading edge response** is a key part of the Kirchhoff approximation response and in fact at high frequencies is part of the exact response of flaws that is often the dominant response. Shown is the leading edge pitch-catch, same mode response of a general complex-shaped convex flaw (void or inclusion). The leading edge response amplitude depends on the plane wave reflection coefficient and the Gaussian curvature of the flaw at a point on the surface called the **specular point** where the unit normal to the surface and the incident and scattered wave directions are related through Snell's law. For cases where there are multiple specular points then all such points contribute to the response.

Flaw Scattering-Inclusion

Leading Edge Response - General Convex Inclusion
(scattered mode same as incident mode)



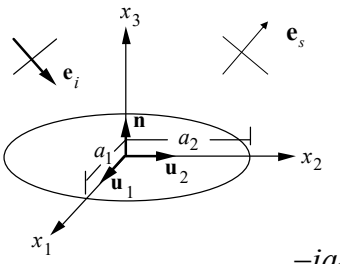
In the time domain $A(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} A(\omega)\exp(-i\omega t)\,d\omega$

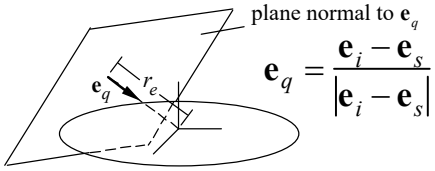
$$A(\mathbf{e}_i;\mathbf{e}_s,t) = R_{12}\frac{\sqrt{R_1R_2}}{2}\delta\left(t+|\mathbf{e}_i-\mathbf{e}_s|\,d/c\right)$$

If we transform the leading edge response into the time domain, we see a delta function.

Flaw Scattering-Crack

Kirchhoff Approximation – Flat Elliptical Cracks





$$\mathbf{e}_q = \frac{\mathbf{e}_i - \mathbf{e}_s}{|\mathbf{e}_i - \mathbf{e}_s|}$$

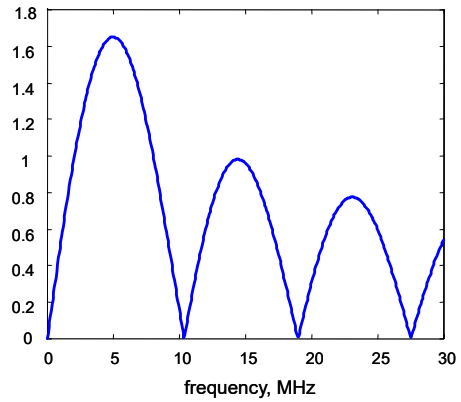
$$A(\mathbf{e}_i;\mathbf{e}_s) = \frac{-ia_1a_2(\mathbf{e}_i \cdot \mathbf{n})}{|\mathbf{e}_i - \mathbf{e}_s| r_e} J_1\left(k|\mathbf{e}_i - \mathbf{e}_s| r_e\right)$$

$$r_e = \sqrt{a_1^2(\mathbf{e}_q \cdot \mathbf{u}_1)^2 + a_2^2(\mathbf{e}_q \cdot \mathbf{u}_2)^2}$$

We can also obtain some solutions for cracks in the Kirchhoff approximation. Here is the pitch-catch response of a flat elliptical crack (fluid model) which involves a **Bessel function, J_1**.

Flaw Scattering-Crack

General pitch-catch crack response-circular crack

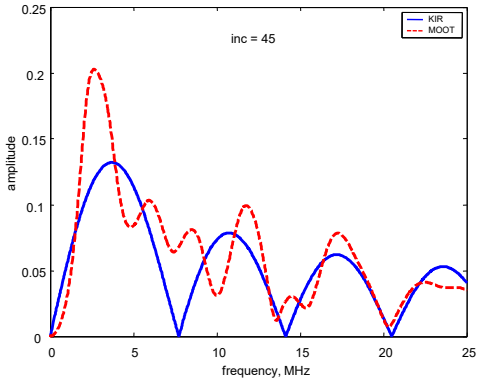


The magnitude of the P-wave pulse-echo far-field scattering amplitude versus frequency for a 1 mm radius circular crack in steel with an angle of incidence of from the crack normal.

If we plot the pitch-catch response in the frequency domain we see a series of decreasing peaks and nulls in the response. These oscillations, we will see, arise because of the existence of so-called flash-points in the time domain response.

Flaw Scattering-Crack

Comparison of the Kirchoff approximation and MOOT for a 0.381 mm radius circular crack at an incident angle of 45 degrees (pulse-echo)

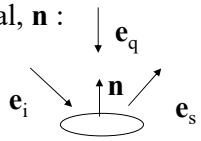


There is a more exact numerical approach to solving scattering problems called MOOT (for the **Method Of Optimal Truncation**) that can be compared with the Kirchhoff approximation result. The MOOT results show the overall behavior of the Kirchhoff approximation result but contain additional oscillations that are due to the interference of other waves that are present besides the flash points.

Flaw Scattering-Crack

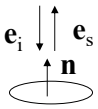
When $\mathbf{e}_q$ is parallel to the crack normal, $\mathbf{n}$:

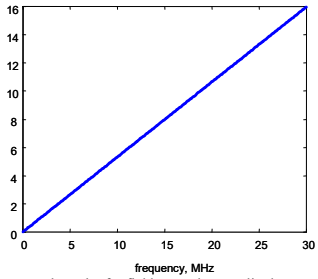
$$A(\mathbf{e}_i;\mathbf{e}_s)=\frac{-ik(\mathbf{e}_i\cdot\mathbf{n})a_1a_2}{2}$$



Special case - pulse echo:

$$A(\mathbf{e}_i;-\mathbf{e}_i)=\frac{ika_1a_2}{2}$$



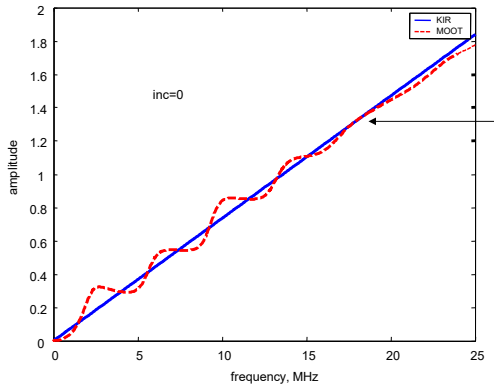


The magnitude of the P-wave pulse-echo far-field scattering amplitude versus frequency for a 1 mm radius circular crack in steel at normal incidence.

In certain cases such as normal incidence (for pulse-echo) the Kirchhoff response is quite different, becoming a linearly increasing function of frequency. We will see this arises because the time-domain response becomes a doublet (derivative of a delta function) in the time domain in this case.

Flaw Scattering-Crack

Comparison of the Kirchoff approximation and MOOT for a 0.381 mm radius crack at normal incidence (pulse-echo)



numerical errors in MOOT

A MOOT solution also shows a general linearly increasing behavior but with some small oscillations that decay at higher frequencies. However, a more careful numerical solution revealed that the oscillations actually continue at these higher frequencies. This MOOT solution did not keep a sufficient number of terms in this solution.

Flaw Scattering-Crack

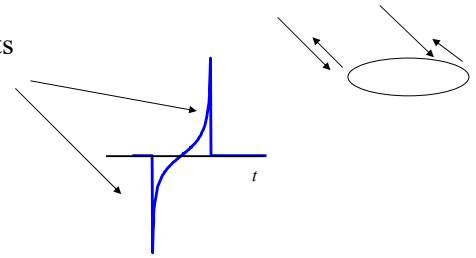
Inverting these results into the time domain

$$\mathbf{e}_i \cdot \mathbf{n} \neq 0$$

$$A(t) \equiv A(\mathbf{e}_i; \mathbf{e}_s, t) = \begin{cases} -\dfrac{a_1 a_2 c (\mathbf{e}_i \cdot \mathbf{n})}{\pi |\mathbf{e}_i - \mathbf{e}_s|^2 r_e^2} \dfrac{t}{\sqrt{(|\mathbf{e}_i - \mathbf{e}_s| r_e / c)^2 - t^2}} & |t| < |\mathbf{e}_i - \mathbf{e}_s| \dfrac{r_e}{c} \\ 0 \quad \textit{otherwise} \end{cases}$$

Example: pulse-echo

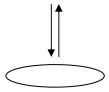
flash points



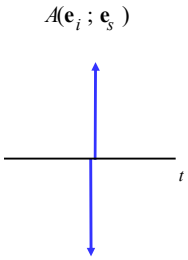
If we invert the crack Kirchhoff response into the time domain we find an antisymmetric pair of pulses called **flash points**, that arise in pulse-echo, for example, when the incident wave first touched the flaw edge and last touches the edge.

Flaw Scattering-Crack

At normal incidence, for pulse-echo



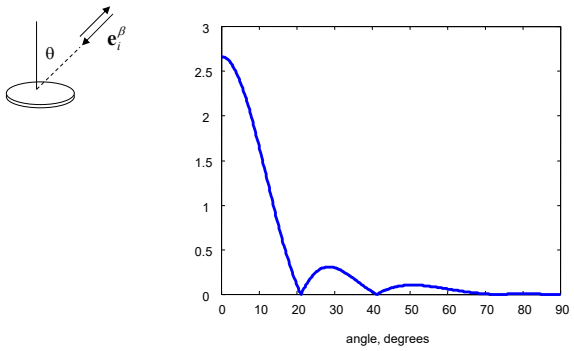
$$A(\mathbf{e}_i;-\mathbf{e}_i,t)=\frac{-a_1a_2}{2c}\frac{d\delta(t)}{dt}$$



In the pulse-echo response at normal incidence, we obtain instead a doublet response (derivative of a delta function) in the time domain that has the linearly increasing frequency response we saw previously.

Flaw Scattering-Crack

A planar crack is a very “specular” reflector

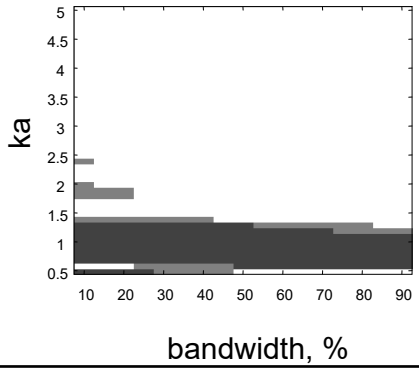


At a given frequency the crack behaves as a very specular reflector, i.e. its response is large in pulse-echo, only when the incident wave direction is normal to the face of the crack. However, real crack responses come from a range of frequencies so that we will show through some numerical studies that this behavior at a single frequency is misleading and the Kirchhoff approximation can often be used at a wide range of angles.

Flaw Scattering-Crack

Although the Kirchhoff approximation is formally a high frequency approximation (ka >>1), numerical tests have shown it capable of producing good agreement (<1 dB difference) for the peak-to-peak time domain response of a circular crack at normal incidence down to ka =1.5 provided the bandwidth is sufficient

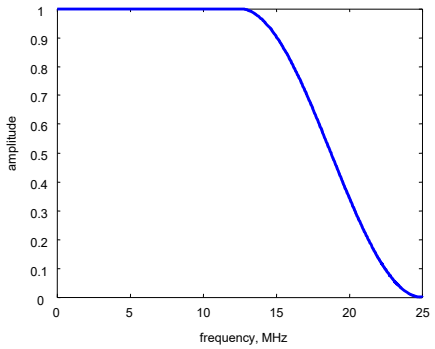
white: differences <1 dB
gray: >1 dB but < 1.5 dB
black: > 1.5 dB



For example, numerical studies have shown that the peak-to-peak amplitude of the crack response is in good agreement with more exact solution to frequencies as low as ka = 1.5 provided the bandwidth is sufficient.

Flaw Scattering-Crack

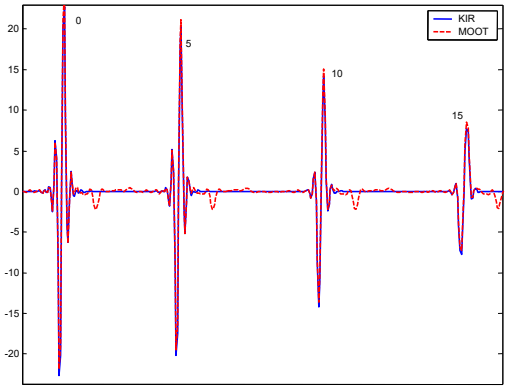
Comparison of synthesized waveforms scattered from a 0.381 mm radius crack using the Kirchhoff approximation and MOOT using frequencies from 0-25 MHz approximately (pulse echo)



We will compare the Kirchhoff and MOOT (considered here to be the “exact” solution) results over a wide band of frequencies and invert these frequency domain results into the time domain so we can examine the scattered waveforms. The MOOT results are left unaltered from zero to 15 MHz, then are smoothly tapered to zero at 25 MHz with the filter shown. This filter generates a wide band response with little ringing in the time domain that can be compared with the Kirchhoff results.

Flaw Scattering-Crack

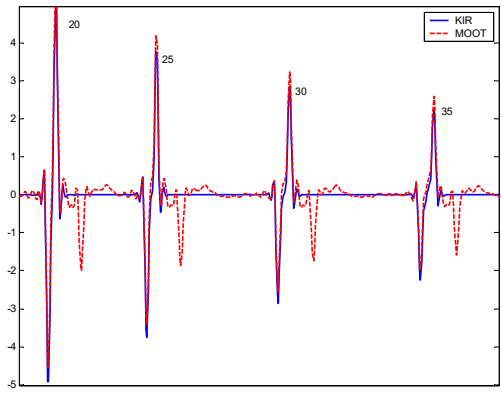
0-15 degrees from crack normal



This shows the Kirchhoff (blue) and MOOT pulse-echo results of a circular crack for angles of 0, 5, 10, and 15 degrees from the normal, respectively (all plotted on the same axis). Generally, we only see a bandlimited doublet response, which is predicted to have the same form for both theories, although the MOOT also contains some small later arriving waves.

Flaw Scattering-Crack

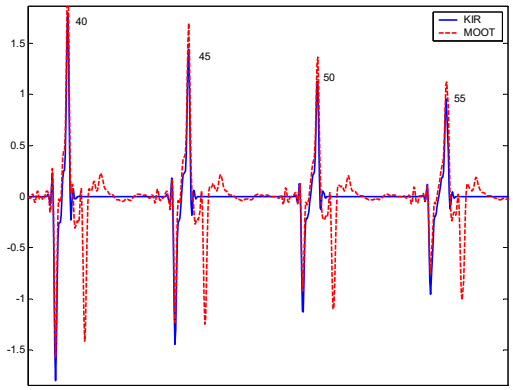
20-35 degrees from crack normal



This comparison is for angles of 20, 25 30 and 35 degrees from the crack normal, where we see the flash points forming in the response but where now MOOT also predicts some larger later arriving waves.

Flaw Scattering-Crack

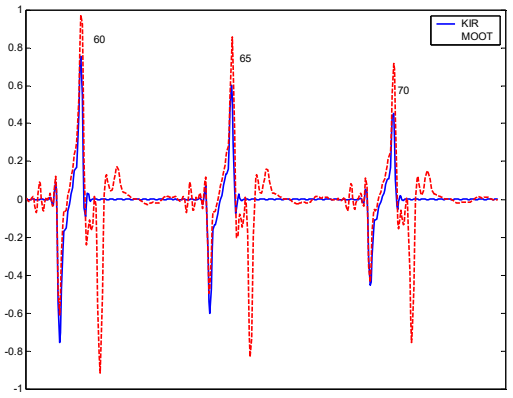
40-55 degrees from crack normal



This continues the study for angles of 40, 45, 50, and 55 degrees from the normal. It is seen the flash points remain the dominant part of the response and that the two theories agree well for those flashpoints (although the second flashpoint is becoming somewhat smaller in the MOOT response from that of the completely antisymmetrical Kirchhoff result) , but that by 55 degrees the later arriving waves have became as large as the first flashpoint.

Flaw Scattering-Crack

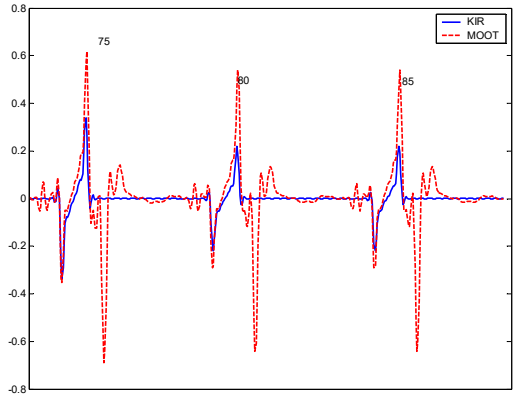
60-70 degrees from crack normal



At 60, 65 and 75 degrees the first flashpoint responses of the two theories still agree quite well but the second flashpoint of the MOOT result is now much smaller than the first flashpoint and the later arriving waves are now larger than the flashpoint responses.

Flaw Scattering-Crack

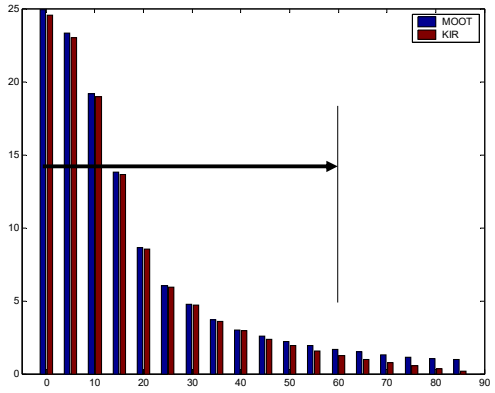
75-85 degrees from crack normal



At the extreme angles of 75, 80, and 85 degrees the general behavior seen on the previous slide continues. If we compare the amplitudes here with the normal incidence case we see the amplitudes here are much smaller, so the crack is indeed rather specular in nature but the Kirchhoff approximation remains valid over a wide range of angles and is not limited in validity to near normal incidence.

Flaw Scattering-Crack

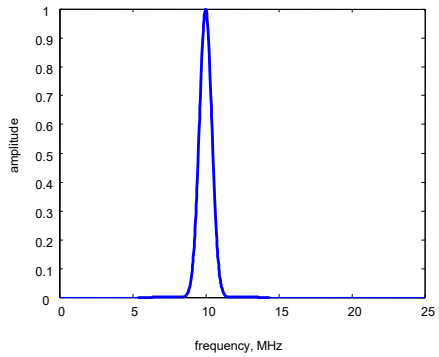
Comparison of peak-to-peak values of MOOT and Kirchhoff versus angle of incidence for the 0.381mm radius crack (pulse-echo)
The arrow shows where agreement is within 1 dB



This slide shows that the peak-to-peak amplitudes of the Kirchhoff and MOOT solutions agree, for these 0-25 MHz bandwidth responses out to about 60 degrees.

Flaw Scattering-Crack

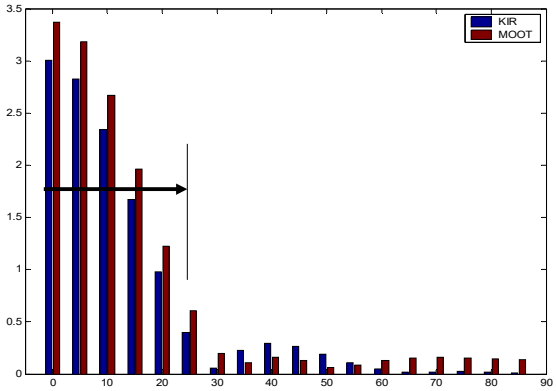
Now consider the scattering amplitude with narrow band Gaussian window
The central frequency is 10 MHz, and the bandwidth is 1 MHz
Radius of the crack a = 0.381mm



In contrast, we now band limit the frequency domain responses to a very narrow bandwidth, as shown.

Flaw Scattering-Crack

the range of angles where the agreement is good is significantly reduced



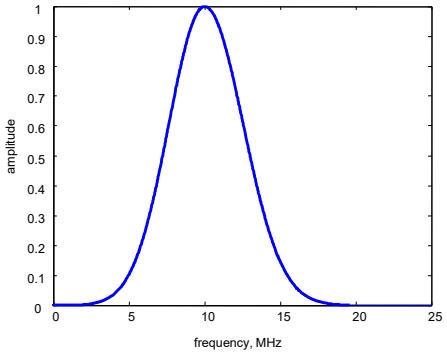
In this narrow bandwidth case the range of angles of agreement is substantially smaller.

Flaw Scattering-Crack

Now consider the scattering amplitude with wider band Gaussian window

The central frequency is 10 MHz, and the bandwidth is 6 MHz

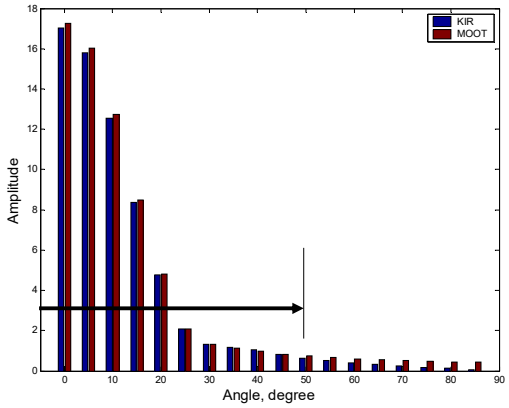
Radius of the crack a = 0.381mm



Now, expand the bandwidth a bit.

Flaw Scattering-Crack

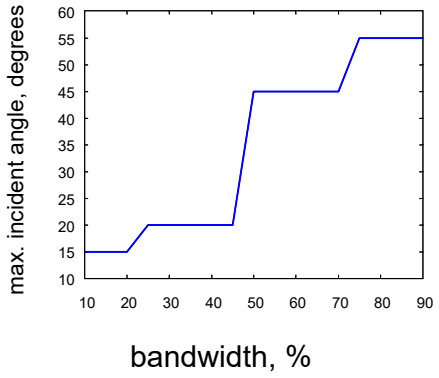
The range of excellent agreement now is back to angles as great as 45 degrees or more



We see the range of angles with good agreement is now substantially larger.

Flaw Scattering-Crack

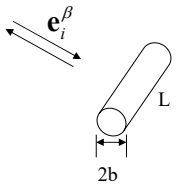
Maximum incident angle where the peak-to-peak time domain agreement between the Kirchhoff approximation and the exact solution for a circular crack is less than 1 dB for ka =5.0 (other ka values shown same trend).



This curve summarizes the effects of bandwidth on the maximum angle at which the Kirchhoff and MOOT results are in good agreement. We see that bandwidth indeed plays a key role in where the Kirchhoff approximation is valid. This is important since in practice the flash point signals are often used in NDE inspections for important tasks such as sizing a crack.

Flaw Scattering-SDH

Kirchhoff approximation for pulse-echo scattering of a side-drilled hole (incident direction in plane perpendicular to the hole axis)



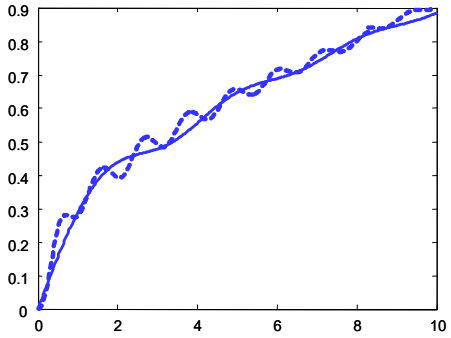
$$A\left(\mathbf{e}_i^{\beta};-\mathbf{e}_i^{\beta}\right)=\frac{\left(k_\beta b\right)L}{2}\left[J_1\left(2k_\beta b\right)-iS_1\left(2k_\beta b\right)\right]+\frac{i\left(k_\beta b\right)L}{\pi}$$

Bessel function Struve function

In addition to spherical voids and circular or elliptical flat cracks, we can get an explicit result for the pulse-echo scattering amplitude of a side-drilled hole (for either incident/scattered P-waves or S-waves). This type of scatterer is often used in calibration studies. We see the response is in terms of a Bessel function, J_1 , and a Struve function, S_1 .

Flaw Scattering-SDH

Comparison with exact 2-D separation of variables solution for P-waves (pulse-echo)

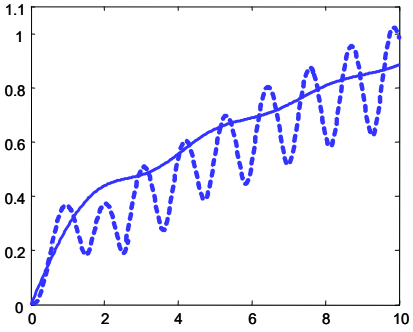


The three-dimensional normalized pulse-echo P-wave scattering amplitude versus normalized wave number for a side drilled hole in the Kirchhoff approximation (solid line) and from the exact two-dimensional separation of variables solution (dashed line).

If we compare the Kirchhoff response (solid line) with an “exact” separation of variables solution (dashed line) for P-waves we see good agreement but as in other cases there are more oscillations in the exact result, likely coming from the interference of waves that are not included in the Kirchhoff approximation.

Flaw Scattering-SDH

Comparison with exact 2-D separation of variables solution for S-waves (pulse-echo_)



The three-dimensional normalized pulse-echo SV-wave scattering amplitude versus normalized wave number for a side drilled hole in the Kirchhoff approximation (solid line) and from the exact two-dimensional separation of variables solution (dashed line).

In contrast, for S-waves incident on the side-drilled hole there are more severe oscillations in the separation of variables solution.

Flaw Scattering

Kirchhoff approximation - Summary

For volumetric flaws -the Kirchhoff approximation properly models the leading edge signal as long as ka >1 approximately but does not model other waves (creeping waves, etc.)

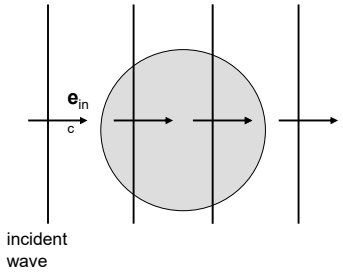
For cracks – the Kirchhoff approximation models the flash point signals properly in pulse-echo as long as ka > 1 approximately and the incident angle is less than about 50 degrees for wide band responses. For narrow band responses this angle is considerably reduced to as little as 15-20 degrees.

In summary, the Kirchhoff approximation , while it does not capture all the aspects of the flaw response, it does model very well the leading edge signals of volumetric flaws and the flashpoint signals of cracks and those signals are often the dominant signals seen and used in NDE tests. In fact, we have shown that it is the leading edge responses of volumetric flaws (and the flashpoints, for cracks) that are primarily responsible for generating the flaw images seen in popular flaw imaging methods such as SAFT (synthetic aperture focusing technique) and Full-Matrix Capture imaging.

Flaw Scattering -Inclusion

The Born Approximation

The Born approximation assumes that the material of an inclusion differs little from the host material so that to first order the incident wave passes through the inclusion unchanged (weak scattering, low frequency approximation)



Another approximation that has been used is the Born approximation that assumes the flaw is nearly the same as the host material so that to first order the incident wave is unchanged as it passes through the flaw. It is a weak scattering, low frequency approximation.

Flaw Scattering -Inclusion

The Born approximation generally is developed from a volume integral expression for the scattering amplitude

density difference difference in elastic constants

$$A\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{\alpha}\right)=\frac{-d_q^{\alpha}}{4\pi\rho c_{\alpha}^2}\int_{V_f}\left[\Delta\rho\,\omega^2\tilde{u}_q+ik_{\alpha}e_{sk}^{\alpha}\Delta C_{kqmj}\frac{\partial\tilde{u}_m}{\partial x_j}\right]\exp\left(-ik_{\alpha}\mathbf{x}\cdot\mathbf{e}_s^{\alpha}\right)dV\left(\mathbf{x}\right)$$

fields in the flaw replaced by incident fields

$$\tilde{u}_q=\tilde{u}_q^{incident}\qquad\frac{\partial\tilde{u}_m}{\partial x_j}=\frac{\partial\tilde{u}_m^{incident}}{\partial x_j}$$

The Born approximation is based on a volume integral representation of the scattering amplitude and simply replaces the fields in that volume integral by the incident wave fields, leading to an explicit representation for the scattering amplitude.

Flaw Scattering -Inclusion

For a spherical inclusion in pulse-echo

spherical Bessel function

$$A\left(\mathbf{e}_i^{\beta};-\mathbf{e}_i^{\beta}\right)=-4k_{\beta}^2b^3F\,\frac{j_1\left(2k_{\beta}b\right)}{2k_{\beta}b}$$

where $$F=\frac{1}{2}\left(\frac{\Delta\rho}{\rho}+\frac{\Delta c}{c_{\beta}}\right)$$

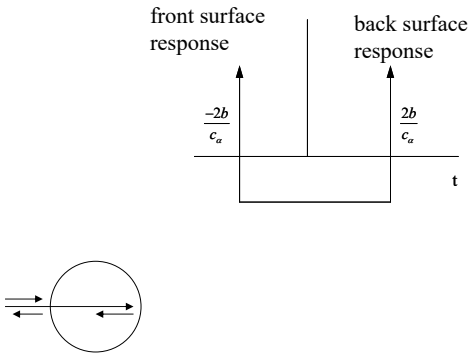
relative density difference

relative wave speed difference

For the pulse-echo response of a spherical inclusion the response is in terms of a spherical Bessel function, j_1 .

Flaw Scattering -Inclusion

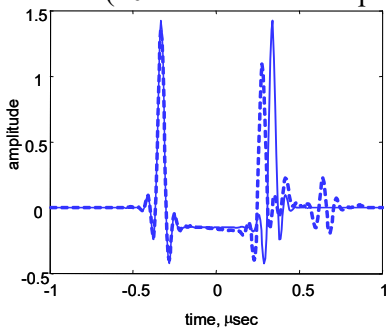
Corresponding time domain impulse response for a spherical inclusion (pulse-echo)



If one inverts the spherical inclusion response into the time domain one finds delta function front and back surface responses and a constant response throughout the flaw in between the front and back surfaces.

Flaw Scattering -Inclusion

Comparison with "exact" separation of variables solution for the pulse-echo response of a spherical inclusion by synthesizing a time-domain response from frequencies ranging from 0-20 MHz approximately for a 1 mm radius inclusion (10 % differences in properties)

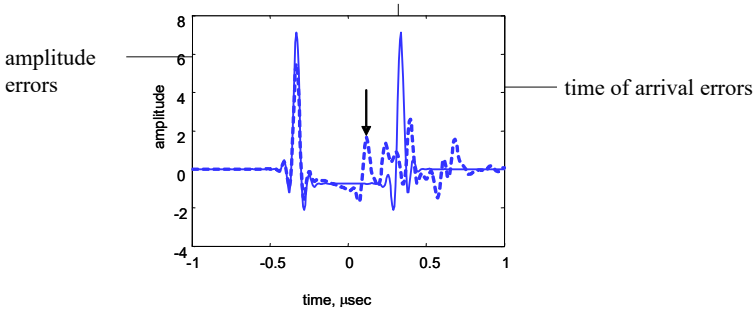


The time domain pulse-echo P-wave response of a 1 mm radius spherical inclusion in steel where the density and compressional wave speed are both ten percent higher than the host steel. Solid line: Born approximation, dashed line: separation of variables solution.

Here is a comparison of the Born approximation with an “exact” separation of variables solution for a weakly scattering spherical inclusion, showing the leading delta function responses are nearly identical and the back surface responses are similar but somewhat displaced. In addition, there are other responses of later arriving waves in the “exact” solution not predicted by the Born approximation.

Flaw Scattering -Inclusion

Comparison with "exact" separation of variables solution for the pulse-echo response of a spherical inclusion by synthesizing a time-domain response from frequencies ranging from 0-20 MHz approximately for a 1 mm radius inclusion (50 % differences in properties)



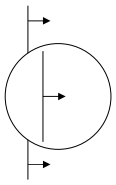
The time domain pulse-echo P-wave response of a 1 mm radius spherical inclusion in steel where the density and compressional wave speed are both fifty percent higher than the host steel.Solid line: Born approximation, dashed line: separation of variables solution.

For a spherical inclusion with properties significantly different from that of the host the front surface signals are very similar but the back surface responses are significantly different in amplitude and time of arrival and there are again other significant waves present not predicted by the Born approximation.

Flaw Scattering -Inclusion

Doubly Distorted Born Approximation

$$A_{Born}\left(\mathbf{e}_i^{\beta};-\mathbf{e}_i^{\beta}\right)=-4k_{\beta}^2b^3F\frac{j_1\left(2k_{\beta}b\right)}{2k_{\beta}b} \qquad F=\frac{1}{2}\left(\frac{\Delta\rho}{\rho}+\frac{\Delta c}{c_{\beta}}\right)$$



replace wave speeds of host material by that of the flaw in the Born approximation

for the pulse-echo response of a spherical inclusion

$$A\left(\mathbf{e}_i^{\beta};-\mathbf{e}_i^{\beta}\right)_{DDBA}=-4k_{f\beta}^2b^3\tilde{F}\frac{j_1\left(2k_{f\beta}b\right)}{2k_{f\beta}b}$$

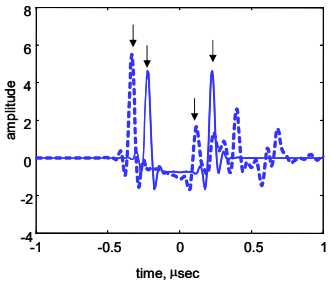
where

$$\tilde{F}=\frac{1}{2}\left(\frac{\Delta\rho}{\rho_f}+\frac{\Delta c}{c_{f\beta}}\right)$$

Since most real flaws found in NDE inspections are not weak scatterers, it would be nice to be able to “fix” the time of arrival and amplitude errors we have just seen. An ad-hoc fix called the doubly distorted Born approximation was developed that simply replaces the wave speed in the F function and the spherical Bessel function by the flaw wave speed rather than the surrounding host wave speed found in the Born approximation. This makes sense since the wave propagating in the flaw travels at the flaw wave speed not that of the host.

Flaw Scattering -Inclusion

Front surface amplitude response is improved, and relative time of arrival of back surface response is correct, but there is an error in absolute time of arrivals (50% differences in properties)

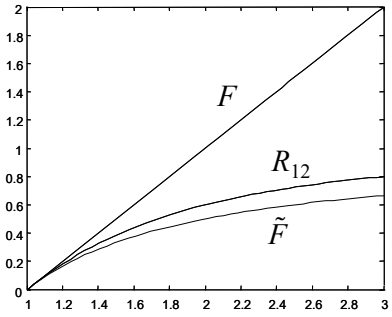


The time domain pulse-echo P-wave response of a 1 mm radius spherical inclusion in steel where the density and compressional wave speed are both fifty percent higher than the host steel. Solid line: Doubly Distorted Born approximation, dashed line: separation of variables solution.

Comparisons of the doubly distorted Born approximation with the separation of variables solution for a strongly scattering spherical inclusion shows that the doubly distorted Born approximation does now get the time separation between the front and back signals correct but the amplitudes of both signals are still different and the front and back signals do not have the correct times of arrival.

Flaw Scattering -Inclusion

Why is the front surface response amplitude improved by the Doubly Distorted Born Approximation ?



$$\frac{\Delta\rho}{\rho} = \frac{\Delta c}{c}$$

Answer: because $\tilde{F} \cong R_{12}^{\beta;\beta}$ (plane wave reflection coefficient)

The amplitude of the front response in the doubly distorted Born approximation is closer to the exact solution than the ordinary Born approximation. The modified F function we see is closer to the exact plane wave reflection coefficient for the flaw so it seems likely that better agreement is the reason for the improvement.

Flaw Scattering -Inclusion

This suggests we apply a phase correction to the doubly distorted Born approximation and replace $\tilde{F}$ by $R_{12}^{\beta;\beta}$ resulting in the modified Born approximation (MBA)

for the pulse-echo response of a spherical inclusion

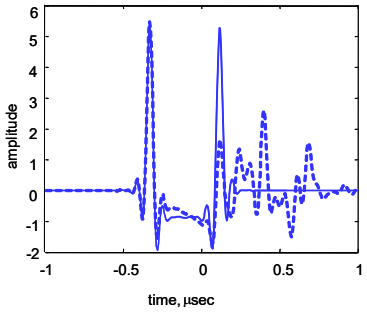
$$A\left(\mathbf{e}_i^{\beta};-\mathbf{e}_i^{\beta}\right)_{MD^2}=-4k_{f\beta}^2b^3R_{12}^{\beta;\beta}\exp\left[2ik_{f\beta}b\left(1-c_{f\beta}/c_{\beta}\right)\right]\frac{j_1\left(2k_{f\beta}b\right)}{2k_{f\beta}b}$$

phase correction

Thus, instead of using the doubly distorted Born approximation, it makes sense instead to replace the F function by the plane wave reflection coefficient and add a phase term to account for the time of arrival errors. This simple modification we have called **the modified Born approximation**.

Flaw Scattering -Inclusion

The MBA
(50 % differences in properties)

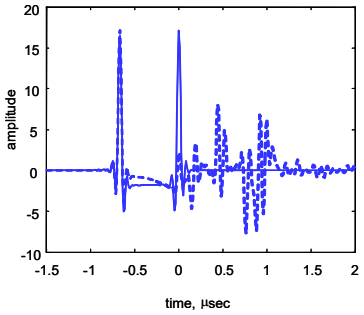


The time domain pulse-echo P-wave response of a 1 mm radius spherical inclusion in steel where the density and compressional wave speed are both fifty percent higher than the host steel.Solid line: MBA approximation, dashed line: separation of variables solution.

Here is now a comparison of the modified Born approximation with the separation of variables solution for a strongly scattering inclusion. We see the font surface responses now match almost perfectly and the time of arrivals of both the front and back surfaces are now correct also. The amplitudes of the back surface responses are still different as the Born approximation always assumes the front and back surface responses are identical and this is simply not the case in practice except in the weakly scattering limit.

Flaw Scattering -Inclusion

The MBA
(100 % differences in properties)



The time domain pulse-echo P-wave response of a 1 mm radius spherical inclusion in steel where the density and compressional wave speed are both one hundred percent higher than the host steel. Solid line: MBA approximation, dashed line: separation of variables solution.

Even at 100 per cent differences between the flaw and host properties the modified Born approximation continues to get at least the front surface signal correct. Note the many late arriving signals in the exact response. There are many other waves present that travel both in the inclusion and around it that are not accounted by the Born approximation.

Flaw Scattering

The Method of Separation of Variables

The sphere and the cylinder are the only two geometries where we can obtain exact separation of variables solutions for elastic wave scattering problems. These are commonly used as "exact" solutions to test more approximate theories and numerical methods.

The method of separation of variables gives us an “exact” method to calculate scattering properties of spherical and cylindrical flaws so it is an important tool for making comparisons with approximate theories, as we have shown.

Flaw Scattering-Void

Example 1: pulse-echo P-wave scattering of a spherical void

$$A\left(\mathbf{e}_i^p;-\mathbf{e}_i^p\right)=\frac{-1}{ik_p}\sum_{n=0}^{\infty}(-1)^n A_n$$

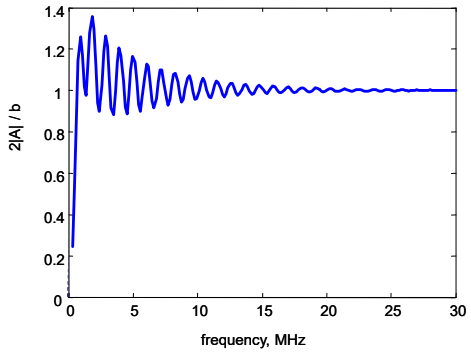
$$A_n=\frac{E_3E_{42}-E_4E_{32}}{E_{31}E_{42}-E_{41}E_{32}}$$

$$E_3=(2n+1)\left\{\left[n^2-n-\left(k_s^2b^2/2\right)\right]j_n\left(k_{pb}\right)+2k_pbj_{n+1}\left(k_pb\right)\right\}$$
$$E_4=(2n+1)\left\{(n-1)\,j_n\left(k_{pb}\right)-k_pbj_{n+1}\left(k_pb\right)\right\}$$
$$E_{31}=\left[n^2-n-\left(k_s^2b^2/2\right)\right]h_n^{(1)}\left(k_pb\right)+2k_pbh_{n+1}^{(1)}\left(k_pb\right)$$
$$E_{41}=(n-1)h_n^{(1)}\left(k_pb\right)-k_pbh_{n+1}^{(1)}\left(k_pb\right)$$
$$E_{32}=-n(n+1)\left[(n-1)h_n^{(1)}\left(k_sb\right)-k_sbh_{n+1}^{(1)}\left(k_sb\right)\right]$$
$$E_{42}=-\left[n^2-1-\left(k_s^2b^2/2\right)\right]h_n^{(1)}\left(k_sb\right)-k_sbh_{n+1}^{(1)}\left(k_sb\right)$$

The separation of variables solution for the pulse-echo P-wave response for a spherical void is shown here. It involves an infinite sum of special function terms which must be truncated for numerical purposes. Generally more terms are needed at the higher frequencies. It is not difficult to calculate on the order of 100 terms in the sum and this typically is enough to cover the frequencies found in most NDE tests.

Flaw Scattering-Void

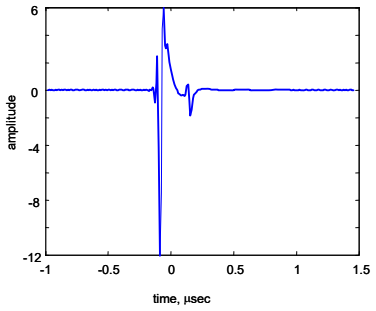
The normalized P-wave pulse-echo scattering amplitude 2A/b for a spherical void of radius b.



Here is the pulse echo response of a spherical void as calculated with the method of separation of variables. It looks similar to the Kirchhoff approximation response. At high frequencies the amplitudes agree but the oscillations are different, especially at low frequencies.

Flaw Scattering-Void

Using this separation of variables solutions at many frequencies to synthesize a P-wave impulse time domain solution (pulse-echo)



The time-domain pulse-echo P-wave response of a 0.5 mm radius spherical void in steel (c_p = 5900 m/s, c_s = 3200 m/sec) obtained by applying a low-pass cosine-squared windowing filter between 10 and 20 MHz to the separation of variables solution and then inverting the result into the time domain with the inverse Fourier transform.

If we use that frequency domain response and calculate the time domain response we see the leading edge response and a response over the lit surface which is not constant as found in the Kirchhoff approximation. There also is a later arriving creeping wave which has traveled around the sphere and is not predicted by The Kirchhoff approximation.

Flaw Scattering-Void

Example 2: pulse-echo SV-wave scattering of a spherical void

$$A\left(\mathbf{e}_i^s;-\mathbf{e}_i^s\right)=\frac{-1}{ik_s}\sum_{n=1}^{\infty}\frac{(-1)^n(2n+1)B_n}{2}$$

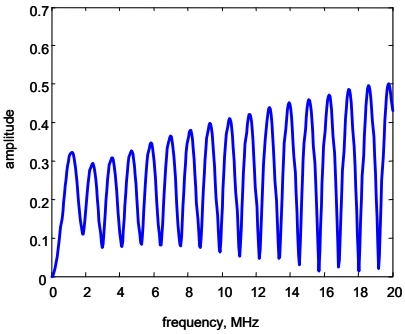
$$B_n=\frac{H_{13}J_{42}-H_{43}J_{12}}{H_{13}H_{42}-H_{43}H_{12}}-\frac{J_{41}}{H_{41}}$$

$$J_{12}=n(n+1)\left[(n-1)j_n(k_sb)-k_sb\,j_{n+1}(k_sb)\right]$$
$$H_{12}=n(n+1)\left[(n-1)h_n^{(1)}(k_sb)-k_sb\,h_{n+1}^{(1)}(k_sb)\right]$$
$$H_{13}=\left[n^2-n-\left(k_s^2b^2/2\right)\right]h_n^{(1)}(k_pb)+2k_pb\,h_{n+1}^{(1)}(k_pb)$$
$$J_{41}=(n-1)j_n(k_sb)-k_sb\,j_{n+1}(k_sb)$$
$$H_{41}=(n-1)h_n^{(1)}(k_sb)-k_sb\,h_{n+1}^{(1)}(k_sb)$$
$$J_{42}=\left[n^2-1-\left(k_s^2b^2/2\right)\right]j_n(k_sb)+k_sb\,j_{n+1}(k_sb)$$
$$H_{42}=\left[n^2-1-\left(k_s^2b^2/2\right)\right]h_n^{(1)}(k_sb)+k_sb\,h_{n+1}^{(1)}(k_sb)$$
$$H_{43}=(n-1)h_n^{(1)}(k_pb)-k_pb\,h_{n+1}^{(1)}(k_pb)$$

Here is the separation of variables pulse-echo solution for an SV-wave incident on a spherical void.

Flaw Scattering-Void

The SV-wave pulse-echo scattering amplitude

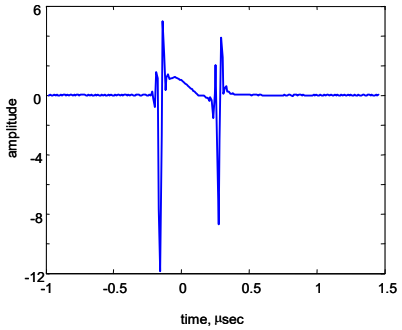


The magnitude of the pulse-echo SV-wave response, , versus frequency for a 0.5 mm radius spherical void in steel (c_p = 5900 m/s, c_s = 3200 m/sec) as calculated by the method of separation of variables

In the frequency domain there now are considerably more oscillations present, suggesting the presence of large, later arriving waves.

Flaw Scattering-Void

Using this separation of variables solutions at many frequencies to synthesize an SV-wave impulse time domain solution (pulse-echo)



The time domain pulse-echo SV-wave response for the same void considered in the P-wave case by applying a low-pass cosine-squared windowing filter between 10 and 20 MHz to the separation of variables solution and then inverting the result into the time domain with the inverse Fourier transform.

Transforming the response into the time domain we see the leading edge response again but now there is a very strong later arriving creeping wave.

Flaw Scattering- SDH

Example 3: pulse-echo P-wave scattering of a cylindrical void

$$\frac{A_{3D}\left(\mathbf{e}_i^p;-\mathbf{e}_i^p\right)}{L}=\frac{i}{2\pi}\sum_{n=0}^{\infty}\left(2-\delta_{0n}\right)\left(-1\right)^n F_n \qquad A_{2D}\left(\omega\right)=\left(\frac{2i\pi}{k_{\alpha 2}}\right)^{1/2}\frac{A_{3D}\left(\omega\right)}{L}$$

$$\delta_{0n}=\begin{cases}1 & n=0\\ 0 & otherwise\end{cases}$$

$$F_n=1+\frac{C_n^{(2)}\left(k_pb\right)C_n^{(1)}\left(k_sb\right)-D_n^{(2)}\left(k_pb\right)D_n^{(1)}\left(k_sb\right)}{C_n^{(1)}\left(k_pb\right)C_n^{(1)}\left(k_sb\right)-D_n^{(1)}\left(k_pb\right)D_n^{(1)}\left(k_sb\right)}$$

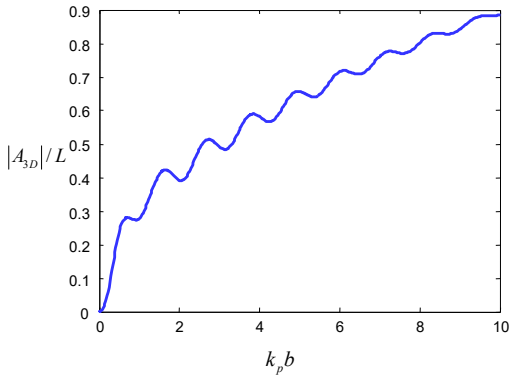
$$C_n^{(i)}\left(x\right)=\left(n^2+n-\left(k_sb\right)^2/2\right)H_n^{(i)}\left(x\right)-\left(2n\,H_n^{(i)}\left(x\right)-x\,H_{n+1}^{(i)}\left(x\right)\right)$$

$$D_n^{(i)}\left(x\right)=n\left(n+1\right)H_n^{(i)}\left(x\right)-n\left(2n\,H_n^{(i)}\left(x\right)-x\,H_{n+1}^{(i)}\left(x\right)\right)$$

We can also write down a separation of variables solution for a P-wave incident on a cylindrical void. This is a 2-D scatterer where we can define its scattering in terms of a 3-D scattering amplitude or a 2-D scattering amplitude, which are simply related.

Flaw Scattering-SDH

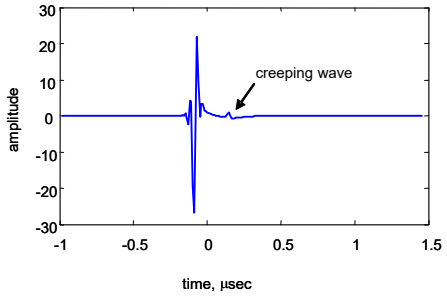
Recall, the scattering amplitude for the pulse-echo P-wave case was:



This gives the response in the frequency domain we saw previously and which was generally in good agreement with the Kirchhoff approximation.

Flaw Scattering- SDH

Using this separation of variables solutions at many frequencies to synthesize a P-wave impulse time domain solution (pulse-echo)



The time-domain pulse-echo P-wave response of a 0.5 mm radius cylindrical void in steel (c_p = 5900 m/s, c_s = 3200 m/sec) obtained by applying a low-pass cosine-squared windowing filter between 10 and 20 MHz to the separation of variables solution and then inverting the result into the time domain with the inverse Fourier transform.

If we invert the result into the time domain we see the leading edge response and a weak creeping wave.

Flaw Scattering - SDH

Example 4: pulse-echo SV-wave scattering of a cylindrical void

$$\frac{A_{3D}\left(\mathbf{e}_i^{sv};-\mathbf{e}_i^{sv}\right)}{L}=\frac{i}{2\pi}\sum_{n=0}^{\infty}\left(2-\delta_{0n}\right)\left(-1\right)^n G_n$$

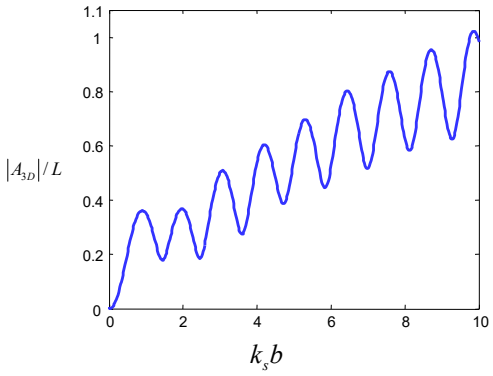
$$\delta_{0n}=\begin{cases}1 & n=0\\ 0 & otherwise\end{cases}$$

$$G_n=1+\frac{C_n^{(2)}\left(k_s b\right)C_n^{(1)}\left(k_p b\right)-D_n^{(2)}\left(k_s b\right)D_n^{(1)}\left(k_p b\right)}{C_n^{(1)}\left(k_p b\right)C_n^{(1)}\left(k_s b\right)-D_n^{(1)}\left(k_p b\right)D_n^{(1)}\left(k_s b\right)}$$

If instead we examine the pulse-echo SV-wave response of the cylindrical void we also can write down the separation of variables solution.

Flaw Scattering -SDH

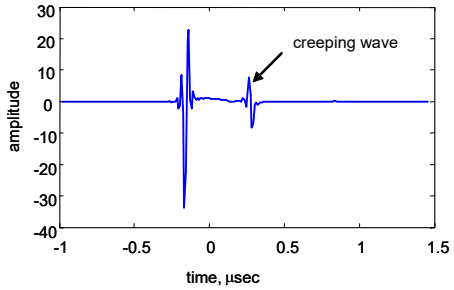
Again, the scattering amplitude for the pulse-echo SV-wave case was:



Recall, this again generally agreed with the Kirchhoff approximation but now there were deeper oscillations in the separation of variables solution.

Flaw Scattering - SDH

Using this separation of variables solutions at many frequencies to synthesize an SV-wave impulse time domain solution (pulse-echo)



The time-domain pulse-echo SV-wave response of a 0.5 mm radius cylindrical void in steel (c_p =5900 m/s, c_s = 3200 m/sec) obtained by applying a low-pass cosine-squared windowing filter between 10 and 20 MHz to the separation of variables solution and then inverting the result into the time domain with the inverse Fourier transform.

Those deeper oscillations are present because the creeping wave response is larger than in the P-wave case.

Flaw Scattering - SDH

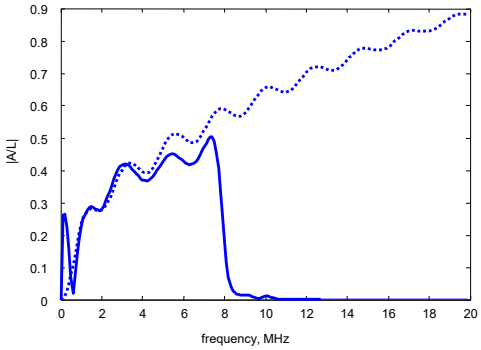
Experimentally determined scattering amplitude by deconvolution (side-drilled hole)

$$V_R(\omega) = G(\omega)\left[\frac{A(\omega)}{L}\right] \qquad G(\omega) = s(\omega)E(\omega)$$

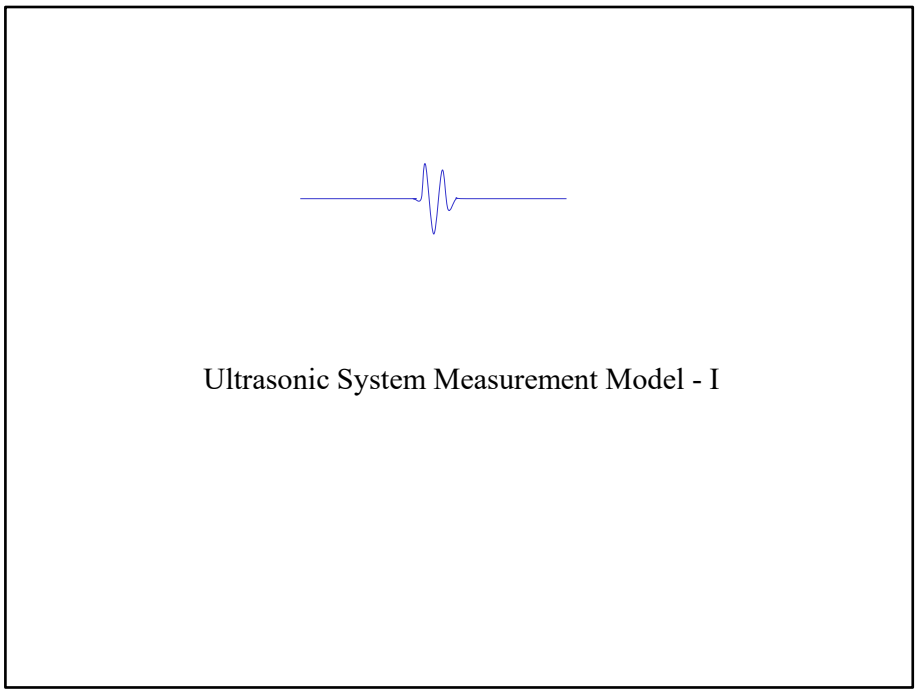
$$E(\omega) = \left[\int_L \hat{V}_0^{(1)}(z,\omega)\hat{V}_0^{(2)}(z,\omega)\,dz\right]\left[\frac{4\pi\rho_2 c_{\alpha 2}}{-ik_{\alpha 2}Z_r^{T;a}}\right]$$

We mentioned earlier that another way to obtain scattering amplitudes is experimentally. For small flaws we can write the received voltage as a product of the scattering amplitude by a factor G which can be obtained through models and measurements. We will not give the details for obtaining G here but we show some of the general parts of that factor here, including the system function and an integral over the length, L, of the inclusion of the fields incident on the flaw, which we can obtain with beam models. Through deconvolution, then we can obtain the scattering amplitude.

Flaw Scattering -SDH



Here is the separation of variables pulse-echo P-wave scattering amplitude response for the cylindrical void and the corresponding scattering amplitude determined experimentally over the bandwidth of a 5 MHz transducer. This comparison shows that one of the challenges of NDE testing is that inherently we are always seeing flaw responses over a limited range of frequencies so that limitation often makes it difficult to use features of the response predicted by models.



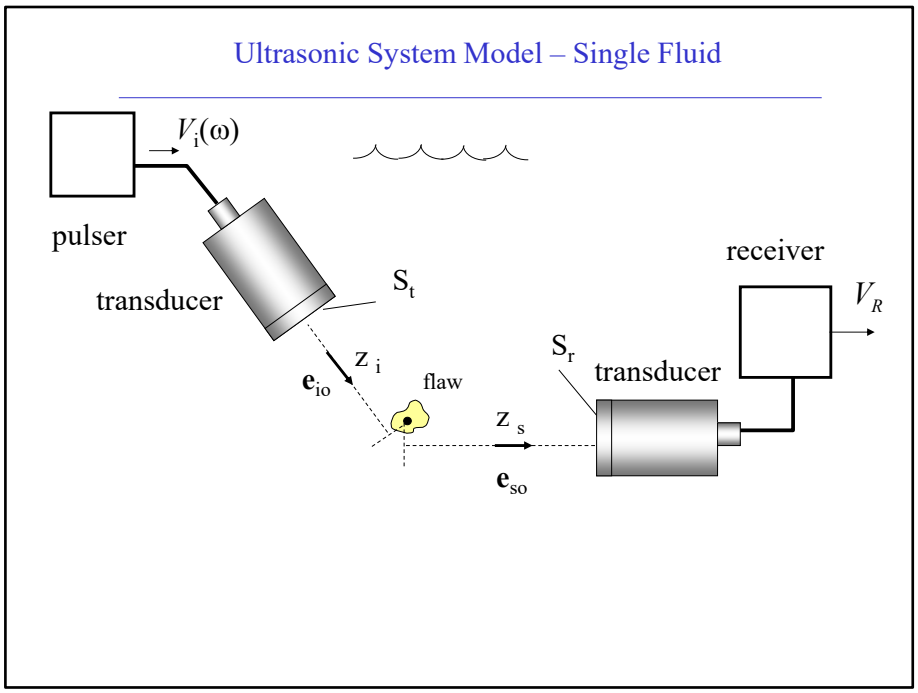
We now want to combine all the elements of an ultrasonic flaw measurement together and generate a complete **ultrasonic measurement model**.

Learning Objectives

Use of a simple fluid model to obtain a complete measurement model

extension to a fluid-solid model

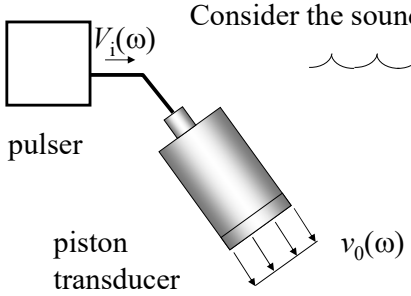
Here, we will use a simple fluid model to develop the ultrasonic measurement model but we will show that this simple model captures all the elements of an immersion NDE measurement.



Here is the configuration we will examine: two transducers examining a scatterer (i.e. “flaw”) in a fluid in a pitch-catch measurement setup.

Ultrasonic System Model – Single Fluid

Consider the sound generation process:



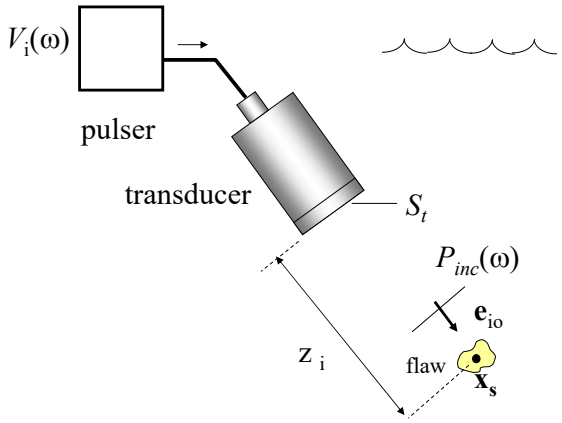
$$F_t(\omega) = \rho c S_t v_0(\omega) = t_G(\omega) V_i(\omega)$$

S_t … area of the transmitting transducer

$t_G(\omega)$ transfer function for pulser, cabling, transducer

We have shown previously that we can write the compressive force generated by the sending piston transducer as a sound generation transfer function multiplied by the Thevenin equivalent voltage put out by the pulser. We saw how we could measure all the elements contained in the sound generation function (pulser impedance, cabling, transducer impedance and sensitivity).

Ultrasonic System Model – Single Fluid

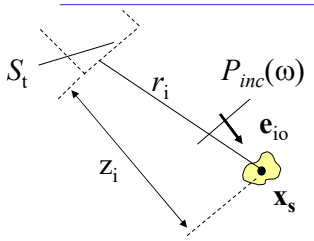


In the paraxial approximation, the beam produces a quasi-plane wave incident on the flaw given by:

$$p(\mathbf{x}_s, \omega) = P_{inc} \exp(ik\mathbf{e}_{i0} \cdot \mathbf{x}_s)$$

In the paraxial approximation the beam of sound produced by the transducer generates a quasi-plane wave incident on the flaw.

Ultrasonic System Model – Single Fluid



$$p(\mathbf{x}_s,\omega) = P_{inc}\exp(ik\mathbf{e}_{i0}\cdot\mathbf{x}_s)$$

$$P_{inc} = \rho c v_0(\omega) C_t(\omega)\exp(ikz_i)$$

$$= \frac{t_G(\omega)V_i(\omega)}{S_t} C_t(\omega)\exp(ikz_i)$$

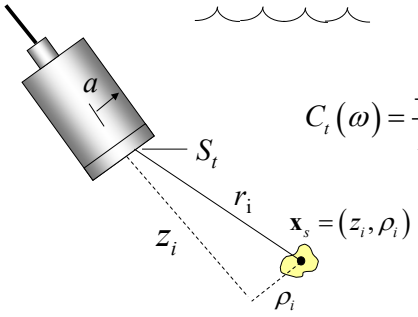
diffraction correction
(deviation from a plane wave)

Note: since

$$p(\mathbf{x}_s,\omega) = \frac{-i\omega\rho v_0}{2\pi}\int_{S_t}\frac{\exp(ikr_i)}{r_i}dS = \rho c v_0 C(\omega)\exp(ikz_i)$$

$$C_t(\omega) = \frac{-ik}{2\pi}\exp(-ikz_i)\int_{S_t}\frac{\exp(ikr_i)}{r_i}dS$$

This quasi plane wave can be written in terms of plane wave traveling to the flaw multiplied by a diffraction correction terms. If we use, for example, a Rayleigh Sommerfeld representation of the beam then we can write that diffraction correction explicitly.



$$C_t(\omega) = \frac{-ik}{2\pi}\exp(-ikz_i)\int_{S_t}\frac{\exp(ikr_i)}{r_i}dS$$

To calculate the diffraction coefficient, however, we can use a multi-Gaussian beam model:

$$p = \rho c v_0 \exp[ikz_i]\sum_{n=1}^{10}\frac{A_n}{[q_1(0)]_n + z_i}\exp\left[\frac{ik\rho_i^2}{2}\frac{1}{[q_1(0)]_n + z_i}\right]$$

$$= \rho c_p v_0 \exp[ik_p z_i] C_t(\omega)$$

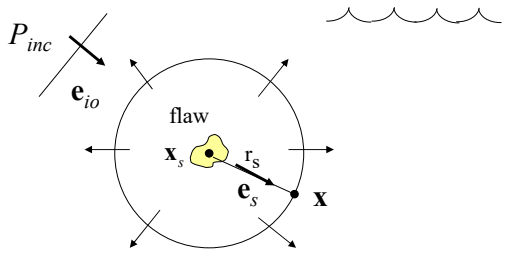
$$C_t(\omega) = \sum_{n=1}^{10}\frac{A_n}{[q_1(0)]_n + z_i}\exp\left[\frac{ik\rho_i^2}{2}\frac{1}{[q_1(0)]_n + z_i}\right]$$

$$[q_1(0)]_n = \frac{-ika^2/2}{B_n}$$

In practice, however, it is easier to use a multi-Gaussian beam model to write the diffraction correction in more explicit terms.

Ultrasonic System Model – Single Fluid

For the waves scattered from the flaw



in the far field

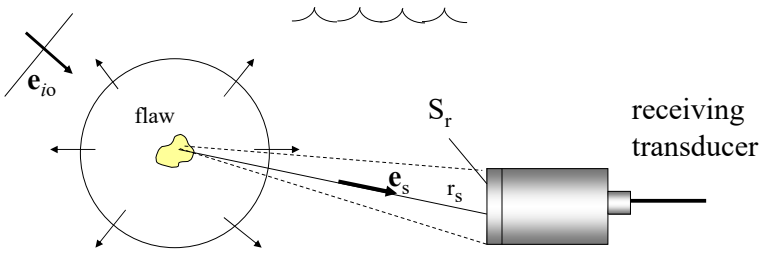
$$p^{scatt}(\mathbf{x},\omega) = P_{inc}\frac{A(\mathbf{e}_{io};\mathbf{e}_s)}{r_s}\exp(ikr_s)$$

$A(\mathbf{e}_{i0};\mathbf{e}_s)$ plane wave far-field scattering amplitude

The scattered wave from the incident quasi-plane wave can be written in terms of the far field scattering amplitude.

Ultrasonic System Model – Single Fluid

Now, consider the scattered waves at the receiving transducer:

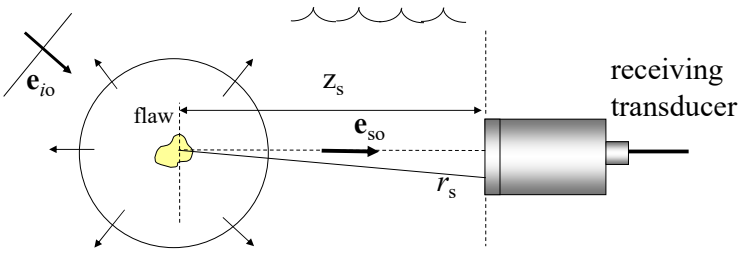


Blocked force:

$$F_B(\omega) = 2\int_{S_r} p^{scatt}\, dS = 2P_{inc}\int_{S_r} A(\mathbf{e}_{io}, \mathbf{e}_s)\frac{\exp(ikr_s)}{r_s}\, dS$$

The scattered waves (which are spreading spherical waves) arrive at the receiving transducer. If we assume plane wave interactions at the receiving transducer, we can just double the scattered pressure field and compute the blocked force at the receiving transducer.

Ultrasonic System Model – Single Fluid

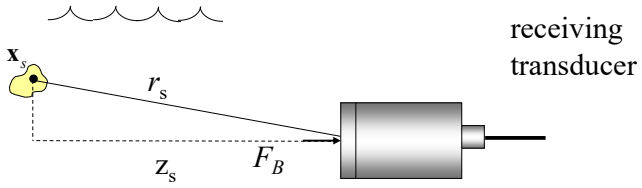


In the paraxial approximation
(on reception) $A(\mathbf{e}_{io}, \mathbf{e}_s) \cong A(\mathbf{e}_{io}, \mathbf{e}_{so})$

and we have $F_B(\omega) = 2P_{inc} A(\mathbf{e}_{io}, \mathbf{e}_{so}) \int_{S_r} \frac{\exp(ikr_s)}{r_s} dS$

Normally, the transducer is far enough from the flaw that the waves arriving are all traveling in essentially a single scattered direction so that the variation of the scattering amplitude over the face of the transducer can be ignored and the scattering amplitude taken out of the integral over the receiving transducer face.

Ultrasonic System Model – Single Fluid



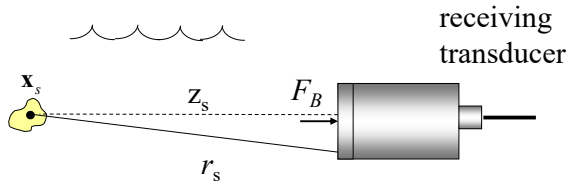
blocked force $$F_B(\omega) = 2P_{inc} A(\mathbf{e}_{io}, \mathbf{e}_{so}) \int_{S_r} \frac{\exp(ikr_s)}{r_s} dS$$

If receiving transducer were acting as a transmitter it would produce a pressure field given by

$$p(\mathbf{x}_s, \omega) = \frac{-i\omega\rho v_0}{2\pi} \int_{S_r} \frac{\exp(ikr_s)}{r_s} dS = \rho c v_0 C_r(\omega) \exp(ikz_s)$$

Thus, the blocked force can be written as a product of the incident amplitude on the flaw times the far field scattering amplitude times a term which is again related to the diffraction correction for the receiving transducer *when it is acting as a transmitter*.

Ultrasonic System Model – Single Fluid



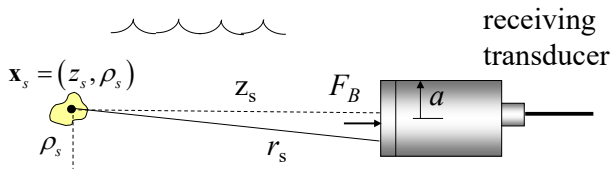
Thus, $$\int_{S_r} \frac{\exp(ikr_s)}{r_s} dS = \frac{2\pi}{-ik} C_r(\omega) \exp(ikz_s)$$

and $$F_B(\omega) = 2P_{inc} A(\mathbf{e}_{io}, \mathbf{e}_{so}) \int_{S_r} \frac{\exp(ikr_s)}{r_s} dS$$

becomes $$= P_{inc} A(\mathbf{e}_{io}, \mathbf{e}_{so}) C_r(\omega) \left[\frac{4\pi}{-ik}\right] \exp(ikz_s)$$

Collecting all these results we have an explicit expression for the blocked force.

Ultrasonic System Model – Single Fluid



$$C_r(\omega) = \frac{-ik}{2\pi}\exp(-ikz_s)\int_{S_r}\frac{\exp(ikr_s)}{r_s}dS$$

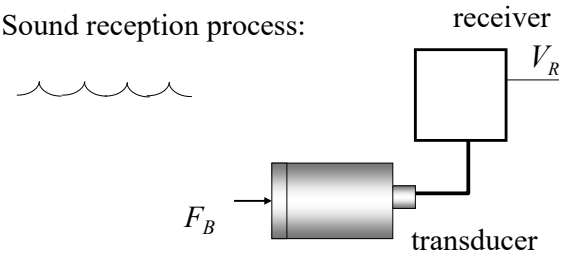
Again, we could compute this diffraction correction with our multi-Gaussian beam model

$$C_r(\omega) = \sum_{n=1}^{10}\frac{A_n}{[q_1(0)]_n + z_s}\exp\left[\frac{ik\rho_s^2}{2}\frac{1}{[q_1(0)]_n + z_s}\right] \qquad [q_1(0)]_n = \frac{-ika^2/2}{B_n}$$

Again it is easier to compute this diffraction correction with a multi-Gaussian beam model.

Ultrasonic System Model – Single Fluid

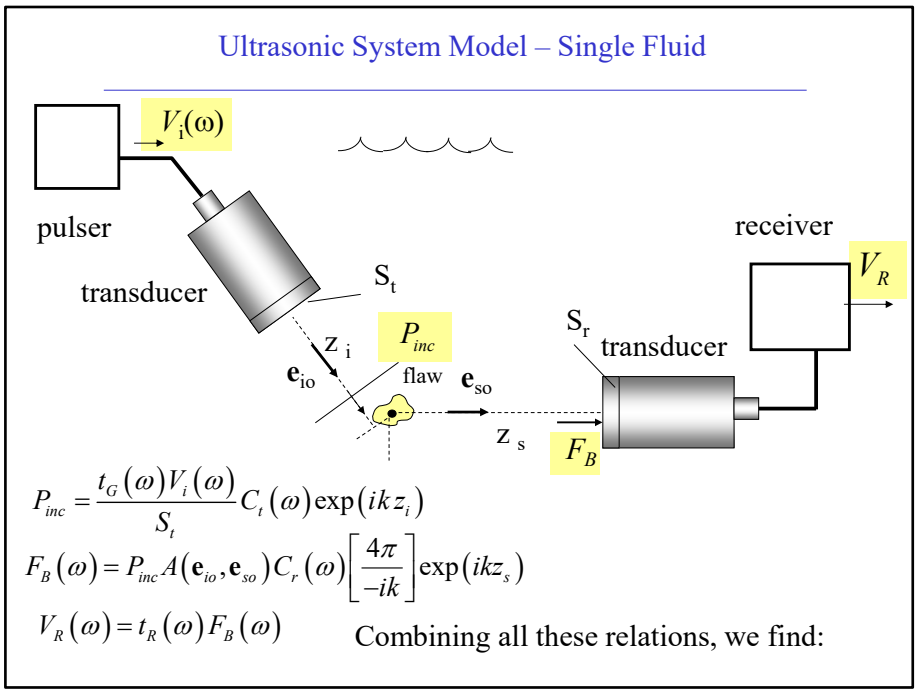
Sound reception process:



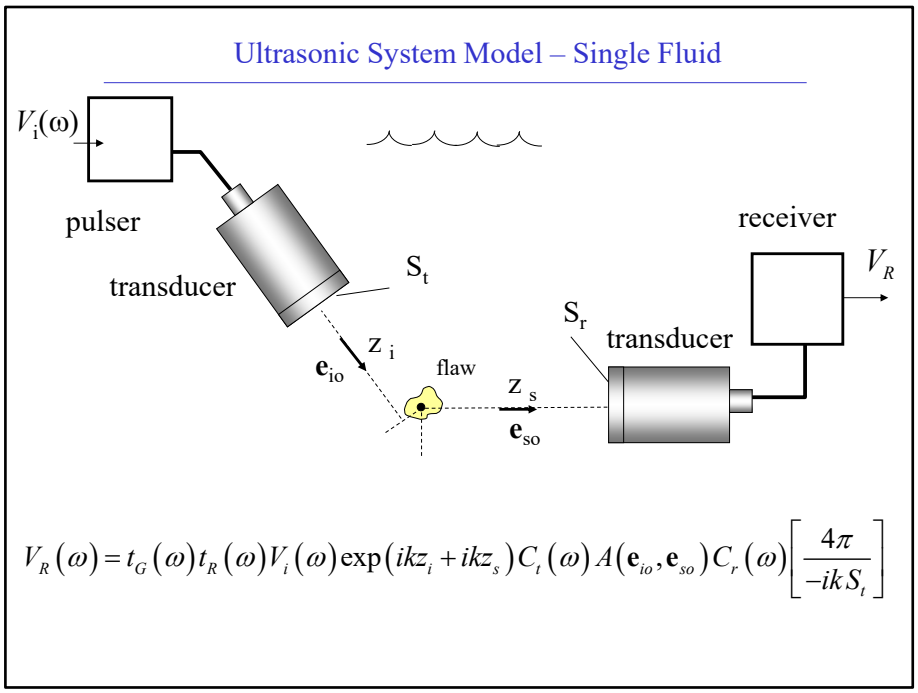
$$V_R(\omega) = t_R(\omega) F_B(\omega)$$

$t_R(\omega)$ transfer function for transducer, cabling, receiver

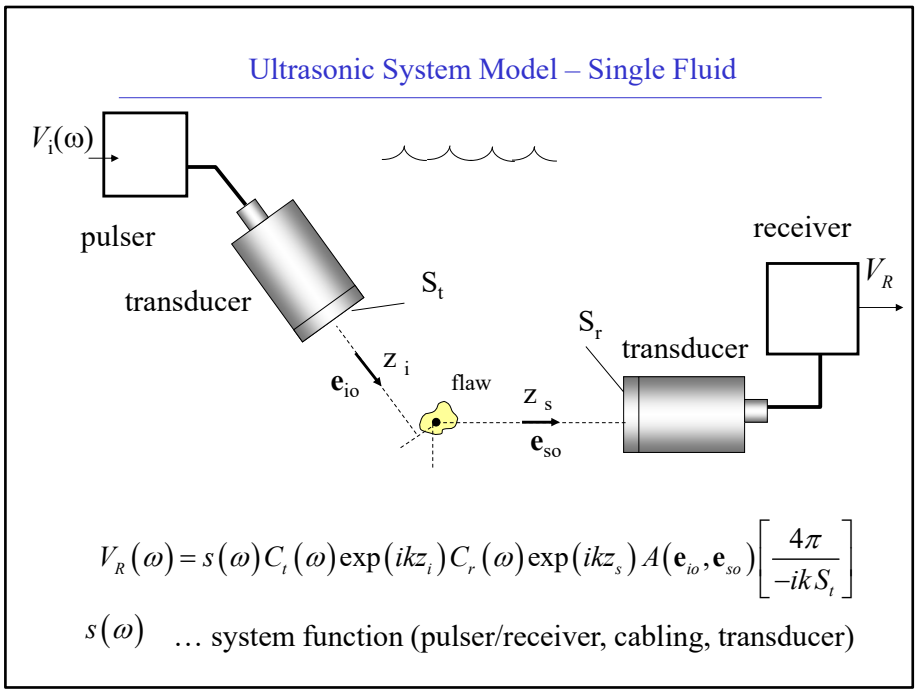
We can relate the received voltage to the blocked force through a sound reception transfer function which accounts for the transducer, cabling and receiver.



We now have terms in our measurement model, as shown, which consider the processes between the pulser and the flaw, the flaw and the receiving transducer, and from the receiving transducer to the measured received voltage.



Combining all the terms we have an (almost) complete measurement model. We say almost complete since we still need to account for wave attenuation, which we will do later.



There are, of course, many measurements needed to obtain all the terms in the measurement model. But this burden can be greatly reduced by replacing many terms in the model by a single system function which can be measured in a calibration setup as we have discussed.

Ultrasonic System Model-Single Fluid

received voltage

$$V_R(\omega) = s(\omega) C_t(\omega) \exp(ikz_i) C_r(\omega) \exp(ikz_s) A(\mathbf{e}_{io}; \mathbf{e}_{so}, \omega) \left[\frac{4\pi}{-ikS_t} \right]$$

- pulser, receiver cabling,transducers
- beam propagation and diffraction effects going from the transmitting transducer to the flaw and from the flaw to the receiving transducer
- flaw scattering
- area of transmitter

When attenuation of the fluid is important, we need to also include terms $\exp\left[-\alpha(\omega) z_i - \alpha(\omega) z_s\right]$

$\alpha(\omega)$ … frequency dependent attenuation

Thus, the measurement model is a product of the system function (which can be measured), a set of terms involving beam propagation and diffraction effects for the sending and receiving transducers (which can be modeled with beam models), and the far field scattering amplitude, which characterizes the flaw response. Finally, there is a remaining term involving a set of constants. To complete this model we can also include wave attenuation terms involving propagation from the sending transducer to the flaw and from the flaw to the receiving transducer.

Ultrasonic System Model-Single Fluid

For pitch-catch

$$V_R(\omega) = s(\omega)\hat{V}_t(\omega)\hat{V}_r(\omega)A(\mathbf{e}_{io};\mathbf{e}_{so},\omega)\left[\frac{4\pi}{-ikS_t}\right]$$

$$\hat{V}_t(\omega) = C_t(\omega)\exp(-\alpha z_i)\exp(ikz_i)$$

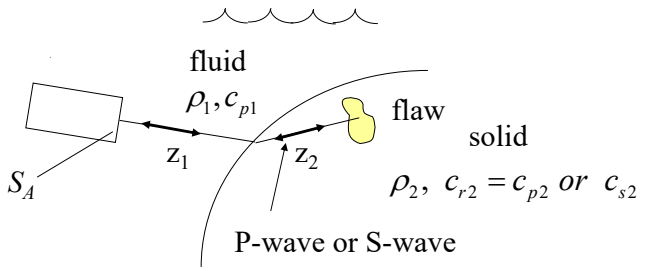
$$\hat{V}_r(\omega) = C_r(\omega)\exp(-\alpha z_s)\exp(ikz_s)$$

For pulse-echo $\hat{V}_t(\omega) = \hat{V}_r(\omega) = \hat{V}(\omega)$

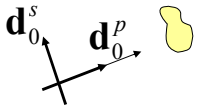
$$V_R(\omega) = s(\omega)\left[\hat{V}(\omega)\right]^2 A(\mathbf{e}_{io};\mathbf{e}_{so},\omega)\left[\frac{4\pi}{-ikS_t}\right]$$

Here we write the pitch-catch measurement model results in a more abbreviated fashion that shows the fields for the sending and receiving transducers explicitly and how the model reduces to a simpler form for the case of a pulse-echo measurement (sending and receiving transducers are identical)

Thompson-Gray Measurement Model – Fluid-Solid (Pulse-Echo)



$$V_R(\omega) = s(\omega)\left[\hat{V}(\omega)\right]^2 A(\omega)\left[\frac{4\pi}{-ik_{r2}S_A}\frac{\rho_2 c_{r2}}{\rho_1 c_{p1}}\right]$$



$$A(\omega) = \left[\mathbf{A}\left(\mathbf{e}_{io}^r; -\mathbf{e}_{io}^r\right)\cdot\left(-\mathbf{d}_0^r\right)\right] \qquad (r = p\ or\ s)$$

We have used a simple fluid model, but if we examine, for example the pulse-echo response of flaw in an actual ultrasonic measurement, a measurement model will be almost identical in form except we can look at P-wave or S-wave responses and the scalar scattering amplitude is a particular component of the vector scattering amplitude that depends on the polarization of the incident waves. The remaining constant coefficient is also slightly different. This measurement model was first obtained in a very similar form by Bruce Thompson and Tim Gray so it is called the **Thompson-Gray measurement model**.

Thompson-Gray Measurement Model – Fluid-Solid (Pulse-Echo)

$$V_R(\omega) = s(\omega)\left[\hat{V}(\omega)\right]^2 A(\omega)\left[\frac{4\pi}{-ik_{r2}S_A}\frac{\rho_2 c_{r2}}{\rho_1 c_{p1}}\right]$$

$(r = p \text{ or } s)$

$$\hat{V}(\omega) = P(\omega)M(\omega)T(\omega)C(\omega)$$

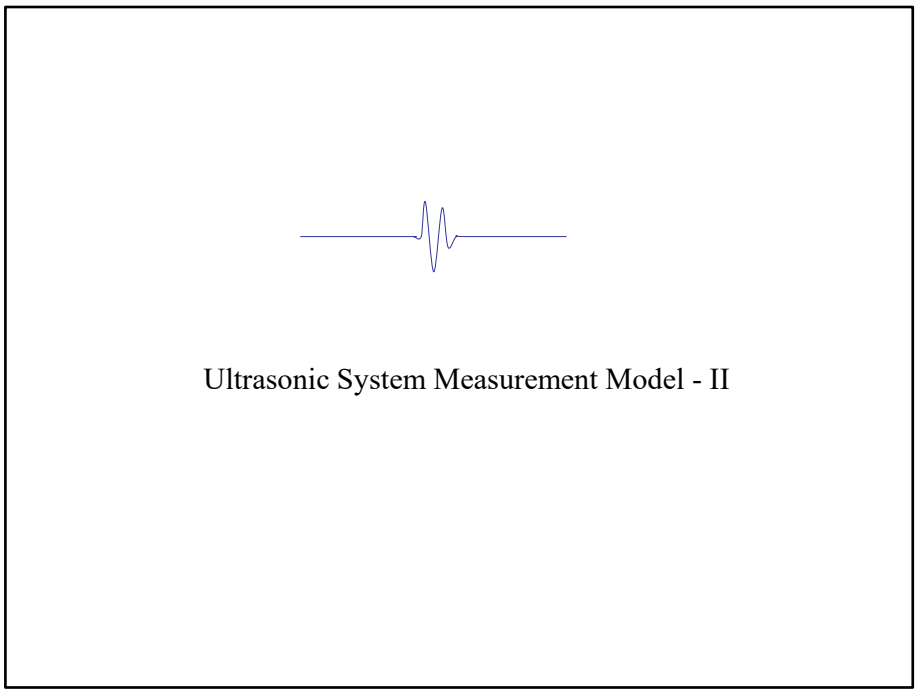
$P(\omega) = \exp\left(ik_{p1}z_1 + ik_{r2}z_2\right)$ … propagation

$M(\omega) = \exp\left(-\alpha_{p1}z_1 - \alpha_r z_2\right)$ … attenuation

$T(\omega) = T_{12}$ … plane wave transmission coefficient (velocity/velocity)

$C(\omega)$ … diffraction coefficient

Here is the Thompson-Gray model with the velocity field expressed more explicitly in terms of propagation, attenuation, transmission, and diffraction terms.



This will be an overview of the various ultrasonic measurement models that have been developed for NDE tests.

Learning Objectives

Definition of a Measurement Model

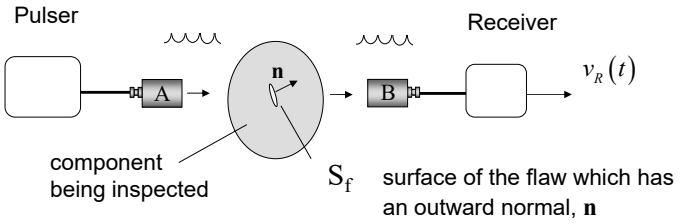
Three Types of Measurement Models
- Auld Model
- Thompson-Gray
- Model for 2D scatterers

Verifying the Modeling Assumptions and Measurements Needed

We will define what an ultrasonic measurement model is and give examples of three types of models that have been developed. We will also discuss the steps we can take to verify the various assumptions made and summarize the measurements needed.

Ultrasonic Measurement Models

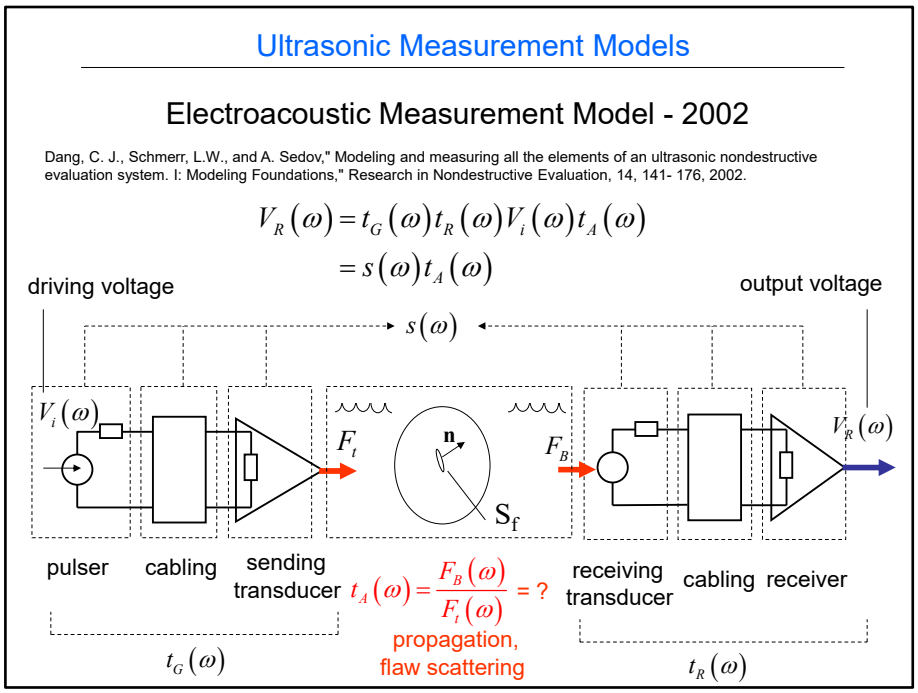
What is an Ultrasonic Measurement Model? A model that expresses the output voltage, $v_R(t)$,received from a flaw in a ultrasonic testing configuration in a form that explicitly gives the modeling or measurement parameters needed to obtain that voltage (immersion shown only as a specific example)



All the measurement models discussed use $V_R(\omega)$, the frequency components of $v_R(t)$

$$V_R(\omega) \overset{FFT}{\leftrightarrow} v_R(t)$$

An ultrasonic measurement model must predict the measured time domain output voltage in a flaw measurement. Our measurement models will all be constructed in the frequency domain but an inverse FFT will yield the desired output voltage signal.



We have seen how the pulser/receiver, cabling, and transducer(s) can all be described in terms of components that can either be modeled or measured. Those aspects of the ultrasonic system can all be combined into a single system function, which we have seen can be measured in a calibration experiment. The remaining part of the system, the acoustic/elastic transfer function, t_A , involves propagation and scattering of the waves present so it is a key element in the measurement model **that must be obtained with models**. A model of an ultrasonic system with all the components shown above described completely was first obtained by Dang et. al. in 2002 and called an **electroacoustic measurement model**

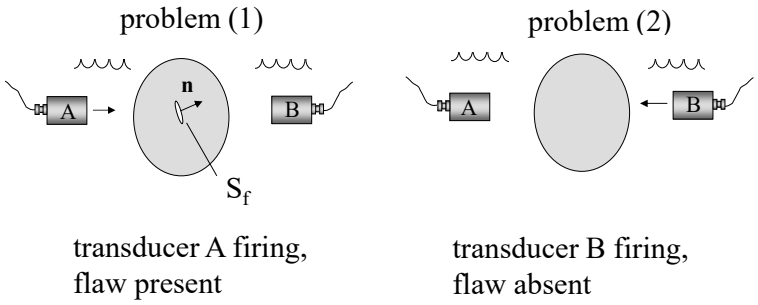
Ultrasonic Measurement Models

Auld Electromechanical Reciprocity Relation

Auld (1979) , using reciprocity relations and the solutions to two specific problems, developed an explicit relation that has been widely used

B. A. Auld, Wave Motion,, Vol. 1, (1979), pp. 3-10.

The two problems:



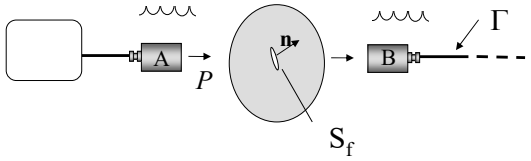
transducer A firing, flaw present

transducer B firing, flaw absent

In 1979 Bert Auld used reciprocity relations and the solution to two specific problems: problem (1), where the flaw was present in the component being inspected (i.e. the problem we want to solve, and problem (2), the same configuration except where the flaw was absent. With these solutions Auld was able to generate a general model of the ultrasonic measurement process. Auld used a contact testing case, rather than the immersion case shown here, in his relations but that difference is not significant.

Ultrasonic Measurement Models

Auld Electromechanical Reciprocity Relation



$$\delta\Gamma = \frac{1}{4P}\int_{S_f}\left(\mathbf{t}^{(1)}\cdot\mathbf{v}^{(2)} - \mathbf{t}^{(2)}\cdot\mathbf{v}^{(2)}\right)\,dS$$

$\delta\Gamma$… change in cable transmission coefficient due to the presence of the flaw

P… power delivered by the transducer

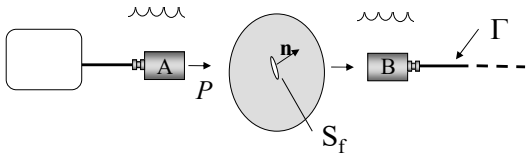
$\mathbf{t}^{(1)}, \mathbf{v}^{(1)}$ stress vector and velocity for problem (1) : A firing and flaw present

$\mathbf{t}^{(2)}, \mathbf{v}^{(2)}$ stress vector and velocity for problem (2): B firing and flaw absent

Auld showed that a change in the signal carried by the cable at the receiver could be expressed in terms of an integral over the flaw surface of the fields in problems (1) and (2).

Ultrasonic Measurement Models

Auld Electromechanical Reciprocity Relation



$$\delta\Gamma = \frac{1}{4P}\int_{S_f}\left(\mathbf{t}^{(1)}\cdot\mathbf{v}^{(2)} - \mathbf{t}^{(2)}\cdot\mathbf{v}^{(2)}\right)\,dS$$

Auld's assumptions:

1) linear, reciprocal electromechanical system (harmonic disturbances)
2) fundamental TEM mode propagating in the cable

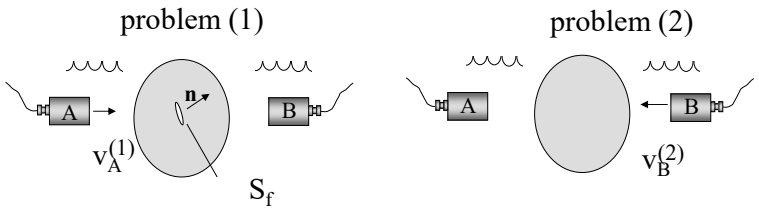
Since we expect $V_R(\omega) \propto \delta\Gamma(\omega)$ Auld's relation "essentially" is an ultrasonic measurement model

Since we expect the change in signal in Auld's model to be directly related to the received voltage signal, Auld's relation is "essentially" a measurement model. It is a very general model since it only assumes that the measurement system is linear and reciprocal and that the electrical disturbances in the cable are propagating TEM mode waves, which are typically the fundamental waves present in coaxial cables.

Ultrasonic Measurement Models

An Alternative Reciprocity Relation

In 1998, Dang, Schmerr, and Sedov obtained a completely explicit model relationship using a purely mechanical reciprocity relationship plus several other relatively weak assumptions



$$V_R(\omega) = \frac{s(\omega)}{Z_r^A(\omega) v_A^{(1)}(\omega) v_B^{(2)}(\omega)} \int_{S_f} \left(\mathbf{t}^{(1)} \cdot \mathbf{v}^{(2)} - \mathbf{t}^{(2)} \cdot \mathbf{v}^{(1)} \right) dS$$

In 1998 Dang et. al. used purely mechanical reciprocity relations to characterize the acoustic/elastic transfer function and arrive at a form that is similar to the ones we have been discussing where the system function appears directly. This is now indeed an ultrasonic measurement model of the Auld form which explicitly expresses the output voltage in terms of the system parameters. Henceforth, we will call this model **Auld's measurement model** in recognition of his foundational contributions.

Ultrasonic Measurement Models

$$V_R(\omega) = \frac{s(\omega)}{Z_r^A(\omega) v_A^{(1)}(\omega) v_B^{(2)}(\omega)} \int_{S_f} \left(\mathbf{t}^{(1)} \cdot \mathbf{v}^{(2)} - \mathbf{t}^{(2)} \cdot \mathbf{v}^{(1)} \right) dS$$

$s(\omega)$ … system function (pulser, cabling, receiver)

$v_A^{(1)}(\omega)$ … average velocity over the face of transducer A in problem (1)

$v_B^{(2)}(\omega)$ … average velocity over the face of transducer B in problem (2)

$Z_r^A(\omega)$ … radiation impedance of transducer A in problem (1)

Here are the basic elements appearing in this form of Auld's measurement model

Ultrasonic Measurement Models

Auld's Measurement Model

$$V_R(\omega) = \frac{s(\omega)}{Z_r^A(\omega) v_A^{(1)}(\omega) v_B^{(2)}(\omega)} \int_{S_f} \left(\mathbf{t}^{(1)} \cdot \mathbf{v}^{(2)} - \mathbf{t}^{(2)} \cdot \mathbf{v}^{(1)} \right) dS$$

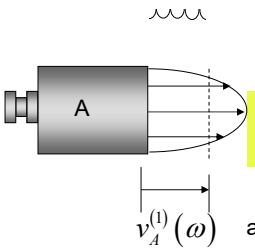
assumptions used to obtain this explicit model

1) linear, reciprocal acoustic and elastic media (harmonic disturbances)

Like Auld's original model, this form relies on very few assumptions. First, there is the assumption that the waves in the system satisfy acoustic/elastic reciprocity.

Ultrasonic Measurement Models

2) we also assumed the output velocity for the transducers can be written in a separable form. For transducer A, for example



ρ_1, c_1 density, wave speed of the fluid

$$v_A^{(1)}(\mathbf{x},\omega) = v_A^{(1)}(\omega) f(\mathbf{x})$$

$v_A^{(1)}(\omega)$ average velocity

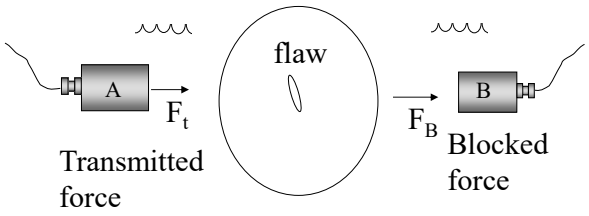
Here, the "forces" are defined in terms of weighted pressure integrals over the transducer face, using the same functions used to describe the velocity. For example, for transducer A

$$F_t(\omega) = \int_{S_A} p(\mathbf{x},\omega) f(\mathbf{x}) dS(\mathbf{x})$$

For a piston transducer $f = 1$

Second, the velocity on the face of the transducers was assumed to be in a separable form where there is a spatial variation terms and a frequency dependent term . This form is satisfied, for example, by a piston transducer, which is a commonly used transducer model.

These two assumptions lead to an explicit expression for the acoustic/elastic transfer function:



$$t_A(\omega) = \frac{F_B(\omega)}{F_t(\omega)} = \frac{1}{Z_r^A(\omega) v_A^{(1)}(\omega) v_B^{(2)}(\omega)} \int_{S_f} \left(\mathbf{t}^{(1)} \cdot \mathbf{v}^{(2)} - \mathbf{t}^{(2)} \cdot \mathbf{v}^{(1)} \right) \, dS$$

With these two assumptions we can obtain an explicit expression for the acoustic/elastic transfer function in terms of an integral over the flaw surface of the fields appearing in problems (1) and (2).

Ultrasonic Measurement Models

3) we also assume that the pulser, receiver, transducers and cabling can be replaced by equivalent linear, time-shift invariant (LTI) systems, i.e.

$$F_t(\omega) = t_G(\omega) V_{in}(\omega)$$
$$V_R(\omega) = t_R(\omega) F_B(\omega)$$

Then, we can write the reciprocity relationship in the form

$$V_R(\omega) = \frac{t_R(\omega) t_G(\omega) V_{in}(\omega)}{Z_r^A(\omega) v_A^{(1)}(\omega) v_B^{(2)}(\omega)} \int_{S_f} \left(\mathbf{t}^{(1)} \cdot \mathbf{v}^{(2)} - \mathbf{t}^{(2)} \cdot \mathbf{v}^{(1)} \right) dS$$

or, finally

$$V_R(\omega) = \frac{s(\omega)}{Z_r^A(\omega) v_A^{(1)}(\omega) v_B^{(2)}(\omega)} \int_{S_f} \left(\mathbf{t}^{(1)} \cdot \mathbf{v}^{(2)} - \mathbf{t}^{(2)} \cdot \mathbf{v}^{(1)} \right) dS$$

We also assume that all the other parts of the system can be represented as LTI systems and that we can lump all those parts into a system function.

Ultrasonic Measurement Models

A General Ultrasonic Measurement Model

$$V_R(\omega) = \frac{s(\omega)}{Z_r^A(\omega) v_A^{(1)}(\omega) v_B^{(2)}(\omega)} \int_{S_f} \left(\mathbf{t}^{(1)} \cdot \mathbf{v}^{(2)} - \mathbf{t}^{(2)} \cdot \mathbf{v}^{(1)} \right) dS$$

assumptions

1) linear, reciprocal acoustic and elastic media (harmonic disturbances)

2) the output velocity for the transducers can be written in a separable form

3) the pulser, receiver, transducers and cabling can be replaced by equivalent linear, time-shift invariant (LTI) systems.

This, then is a general ultrasonic measurement model based on very few assumptions.

Ultrasonic Measurement Models

4) if we also assume that the transducers behave as piston radiators at high frequency then

$$Z_r^A(\omega) = \rho_1 c_1 S_A$$

and our explicit model becomes, finally

$$V_R(\omega) = \frac{s(\omega)}{\rho_1 c_1 S_A v_A^{(1)}(\omega) v_B^{(2)}(\omega)} \int_{S_f} \left(\mathbf{t}^{(1)} \cdot \mathbf{v}^{(2)} - \mathbf{t}^{(2)} \cdot \mathbf{v}^{(1)} \right) dS$$

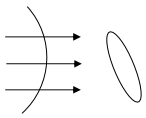
The radiation impedance can normally be taken as its high frequency value, yielding the model shown. This model can be used in essentially the same form for contact testing as well as immersion testing and can include other setups such as angle beam transducers.

Ultrasonic Measurement Models

5) now, assume the incident waves are quasi-plane waves in the vicinity of the flaw, e.g.

$$v_j^{(1)}\Big|_{inc} = V^{(1)}\left(x_1, x_2, x_3, \omega\right) d_j^{(1)} \exp\left(ik_2 \mathbf{e}^{(1)} \cdot \mathbf{x}\right)$$

$$v_j^{(2)} = V^{(2)}\left(x_1, x_2, x_3, \omega\right) d_j^{(2)} \exp\left(ik_2 \mathbf{e}^{(2)} \cdot \mathbf{x}\right)$$



and also define normalized amplitudes of these quasi-plane waves as

$$\hat{V}^{(1)} = \frac{V^{(1)}}{v_A^{(1)}\left(\omega\right)}$$

$$\hat{V}^{(2)} = \frac{V^{(2)}}{v_B^{(2)}\left(\omega\right)}$$

Now, let's assume that the waves incident on the flaw can be describes as quasi-plane waves and normalize the velocity wavefields of these waves by the velocities on the faces of the transducers. Note that we can model these normalized fields with ultrasonic beam models without having to know the amplitudes of the actual velocities present on the transducer faces.

Ultrasonic Measurement Models

Then the form of the measurement model becomes

$$V_R(\omega) = s(\omega)\left[\frac{4\pi\rho_2 c_2}{-ik_2\,\rho_1 c_1 S_A}\right]\int_{S_f} \hat{V}^{(1)}\hat{V}^{(2)}\mathbb{A}\left(\tilde{v}_j^{(1)}, \tilde{\tau}_{ij}^{(1)}\right) dS$$

velocity fields in problems (1),(2) due to unit average velocities on the transducer faces

scattered wave fields from the flaw

Then the form of the measurement model becomes as shown here, where the velocity fields and scattered wave terms all can be described directly with models.

The flaw scattering term is rather complex:

$$\mathcal{A}\left(\tilde{v}_j^{(1)}, \tilde{\tau}_{ij}^{(1)}\right) = \frac{1}{4\pi\rho_2 c_2^2}\left(\tilde{\tau}_{ij}^{(1)} d_i^{(2)} - C_{ijkl} d_k^{(2)}\left(e_l^{(2)} / c_2\right)\tilde{v}_i^{(1)}\right) n_j \exp\left(ik_2 \mathbf{e}^{(2)} \cdot \mathbf{x}\right)$$

$\tau_{ij}^{(1)}, v_j^{(1)}$ stresses and velocity components in problem (1)

$e_j^{(2)}, d_j^{(2)}$ components of the incident wave direction and polarization in problem (2)

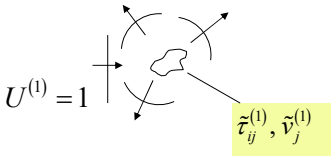
n_j components of the unit outward normal to the flaw surface

C_{ijkl} elastic constants tensor

the normalized fields $\tilde{v}_j^{(1)}, \tilde{\tau}_{ij}^{(1)}$ are defined as velocity and stress in problem (1) due to an incident wave of unit displacement amplitude

$$\tilde{v}_j^{(1)} = \frac{-i\omega v_j^{(1)}}{V^{(1)}}$$

$$\tilde{\tau}_{ij}^{(1)} = \frac{-i\omega \tau_{ij}^{(1)}}{V^{(1)}}$$



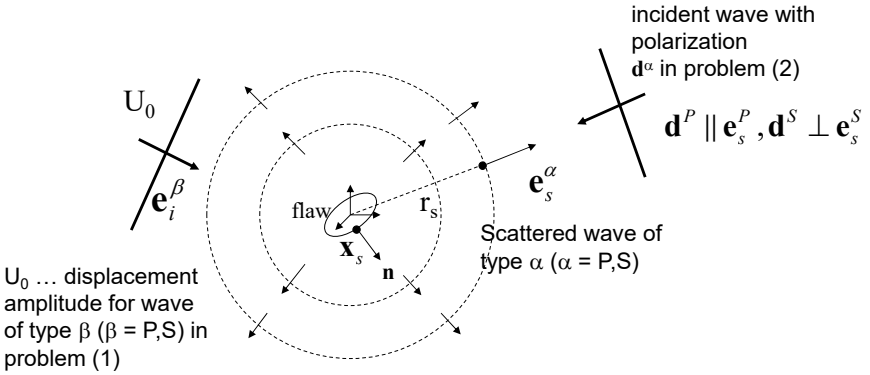
The scattering terms are rather complex functions of the fields on the surface of the flaw. They are computed here as the scattered waves due to an incident wave of unit displacement amplitude so they can be calculated independent of the actual amplitudes of the incident waves, which are contained in the normalized velocity terms. The incident wave is taken to be of unit displacement amplitude since most plane wave scattering amplitudes in elastic solids are defined for such an incident wave.

Ultrasonic Measurement Models

3-D scattering amplitude

$$\mathbf{u}^{scatt}(\mathbf{y},\omega) = U_0 \frac{\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{P}\right)}{r_s}\exp\left(ik_p r_s\right) + U_0 \frac{\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{S}\right)}{r_s}\exp\left(ik_s r_s\right)$$

scattered displacement from a flaw — P-wave — S-wave



Recall the scattered displacements in the far field of the flaw can be described in terms of scattering amplitudes. Also note that if the receiving transducer acts a sending transducer it will generate waves with a polarization vector $\mathbf{d}^P$ or $\mathbf{d}^S$ at the flaw. This polarization vector is important, as we will see.

Ultrasonic Measurement Models

3-D scattering amplitude

$$\mathbf{u}^{scatt}(\mathbf{y},\omega) = U_0 \frac{\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{P}\right)}{r_s}\exp\left(ik_p r_s\right) + U_0 \frac{\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{S}\right)}{r_s}\exp\left(ik_s r_s\right)$$

However, it can be shown that

$$\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{\alpha}\right)\cdot\left(-\mathbf{d}^{\alpha}\right) = \int_{S_f} \mathcal{A}\left(\tilde{v}_j^{(1)},\tilde{\tau}_{ij}^{(1)}\right) dS$$

vector scattering amplitude of waves in problem (1)

polarization vector of incident waves in problem (2)

The component of the scattering amplitude in the negative $\mathbf{d}^P$ or $\mathbf{d}^S$ direction can be shown to be an integral over the surface of the flaw of a scalar function that appears in the measurement model.

Ultrasonic Measurement Models

Measurement Model - Quasi-Plane Waves

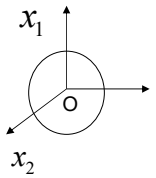
$$V_R(\omega) = s(\omega)\left[\frac{4\pi}{-ik_2 S_A}\frac{\rho_2 c_2}{\rho_1 c_1}\right]\int_{S_f} \hat{V}^{(1)}\hat{V}^{(2)}\mathbb{A}\left(\tilde{v}_j^{(1)}, \tilde{\tau}_{ij}^{(1)}\right) dS$$

$$\mathbf{A}\left(\mathbf{e}_i^{\beta}; \mathbf{e}_s^{\alpha}\right)\cdot\left(-\mathbf{d}^{\alpha}\right) = \int_{S_f} \mathbb{A}\left(\tilde{v}_j^{(1)}, \tilde{\tau}_{ij}^{(1)}\right) dS$$

The beam velocity terms and flaw scattering terms appear separately in this model but they still must be combined before the voltage can be obtained

Thus, the measurement model can be written in terms of the normalized velocity fields and this scalar flaw scattering term.

6) Now, assume a small 3-D flaw, i.e. where the incident waves do not vary significantly in amplitude over the surface of the flaw



$$\hat{V}^{(1)}\left(x_1,x_2,x_3\right)=\hat{V}^{(1)}\left(0,0,0\right)=\hat{V}_0^{(1)}$$
$$\hat{V}^{(2)}\left(x_1,x_2,x_3\right)=\hat{V}^{(1)}\left(0,0,0\right)=\hat{V}_0^{(2)}$$

Then

$$V_R\left(\omega\right)=s\left(\omega\right)\left[\frac{4\pi}{-ik_2S_A}\frac{\rho_2c_2}{\rho_1c_1}\right]\left[\hat{V}_0^{(1)}\hat{V}_0^{(2)}\right]\int_{S_f}\mathbb{A}\left(\tilde{v}_j^{(1)},\tilde{\tau}_{ij}^{(1)}\right)dS$$

and, recall

$$\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{\alpha}\right)\cdot\left(-\mathbf{d}^{\alpha}\right)=\int_{S_f}\mathbb{A}\left(\tilde{v}_j^{(1)},\tilde{\tau}_{ij}^{(1)}\right)dS$$

If we now assume that the flaw is small enough so that the velocity fields do not vary significantly over the flaw surface, then these velocities can be taken out of the integral and the remaining integral is a scalar scattering amplitude term that is the specific component of the vector scattering amplitude defined previously.

Ultrasonic Measurement Models

Measurement Model (Thompson-Gray form)

R. B. Thompson and T. A. Gray, J. Acoust. Soc. Am., Vol. 74, (1983), pp. 140-146.

$$V_R(\omega) = s(\omega)\left[\hat{V}_0^{(1)}\hat{V}_0^{(2)}\right]\left[\mathbf{A}\left(\mathbf{e}_i^{\beta};\mathbf{e}_s^{\alpha}\right)\cdot\left(-\mathbf{d}^{\alpha}\right)\right]\left[\frac{4\pi}{-ik_2S_A}\frac{\rho_2c_2}{\rho_1c_1}\right]$$

beam models

$A\left(\mathbf{e}_i^{\beta},\mathbf{e}_s^{\alpha},\omega\right)$

plane wave far field scattering amplitude component

acoustic/elastic transfer function

$$t_A(\omega) = \left[\hat{V}_0^{(1)}\hat{V}_0^{(2)}\right]\left[A\left(\mathbf{e}_i^{\beta},\mathbf{e}_s^{\alpha},\omega\right)\right]\left[\frac{4\pi}{-ik_2S_A}\frac{\rho_2c_2}{\rho_1c_1}\right]$$

This final form is then in a form due to Thompson and Gray called the **Thompson-Gray measurement model**. A very important feature of this model is that the flaw scattering response is separate from the other terms so that in principle it gives us a way to extract flaw information from the measured ultrasonic signals through deconvolution. In contrast, in the Auld model the scattered waves due to the flaw and transducer(s) wave fields are intermixed. Note that the original derivation of the Thompson-Gray model did not identify the scalar scattering amplitude as an explicit component of the vector far field scattering amplitude, as we have done here.

Ultrasonic Measurement Models

Measurement Model (Thompson-Gray form)

$$V_R(\omega) = s(\omega)\left[\hat{V}_0^{(1)}\hat{V}_0^{(2)}\right]\left[A\left(\mathbf{e}_i^{\beta}, \mathbf{e}_s^{\alpha}, \omega\right)\right]\left[\frac{4\pi}{-ik_2 S_A}\frac{\rho_2 c_2}{\rho_1 c_1}\right]$$

1) linear, reciprocal acoustic and elastic media (harmonic disturbances)

2) the output velocity for the transducer(s) can be written in a separable form

3) the pulser, receiver, transducers and cabling can be replaced by equivalent linear, time-shift invariant (LTI) systems.

4) the transducers are assumed to behave as piston radiators operating at high frequency

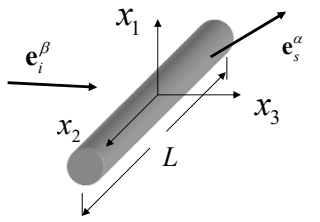
5) the incident waves are quasi-plane waves in the vicinity of the flaw

6) the scatterer is a small 3-D flaw where the incident waves do not vary significantly in amplitude over the surface of the flaw

Here is the Thompson-Gray model and the assumptions that it is based on.

Ultrasonic Measurement Models

6a) If instead, the scatterer is a small cylindrical flaw (e.g. side-drilled hole), with incident and scattered wave directions in the plane of the cross-section then we can assume



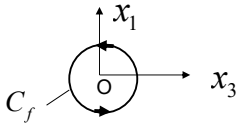
$$\hat{V}^{(1)}(x_1,x_2,x_3)=\hat{V}^{(1)}(0,x_2,0)\equiv\hat{V}_0^{(1)}$$

$$\hat{V}^{(2)}(x_1,x_2,x_3)=\hat{V}^{(1)}(0,x_2,0)\equiv\hat{V}_0^{(2)}$$

$$\mathcal{A}\left(\tilde{v}_j^{(1)},\tilde{\tau}_{ij}^{(1)},\mathbf{x}\right)=\mathcal{A}\left(\tilde{v}_J^{(1)},\tilde{\tau}_{IJ}^{(1)},x_1,x_3\right)\quad (I,J=1,3)\quad \text{only}$$

In this case

$$V_R(\omega)=s(\omega)\left[\frac{4\pi}{-ik_2S_A}\frac{\rho_2c_2}{\rho_1c_1}\right]\int_{length,L}\hat{V}_0^{(1)}\hat{V}_0^{(2)}dx_2\int_{C_f}\mathcal{A}\left(\tilde{v}_J^{(1)},\tilde{\tau}_{IJ}^{(1)}\right)ds$$



For small two-dimensional flaws such as side drilled holes we can again make the assumption that the incident and scattered fields lie basically in a plane perpendicular to the axis of the "flaw" but the incident wave fields will vary along that axis so that we must integrate the incident velocity fields along that axis and the scattering term reduces to a line integral around the cylinder in a plane perpendicular to that axis.

Ultrasonic Measurement Models

3-D scattering amplitude component of a 2-D, cylindrical geometry (end effects neglected) is just

$$A_{3D}\left(\mathbf{e}_i^{\beta},\mathbf{e}_s^{\alpha},\omega\right)=L\int_{C_f}\mathbb{A}\left(\tilde{v}_J^{(1)},\tilde{\tau}_{IJ}^{(1)}\right)ds$$

so that the measurement model can be expressed, finally, as

$$V_R(\omega)=s(\omega)\left[\int_{length,L}\hat{V}_0^{(1)}\hat{V}_0^{(2)}dx_2\right]\frac{A_{3D}\left(\mathbf{e}_i^{\beta},\mathbf{e}_s^{\alpha},\omega\right)}{L}\left[\frac{4\pi}{-ik_2S_A}\frac{\rho_2c_2}{\rho_1c_1}\right]$$

Schmerr, L.W., and A. Sedov, "Modeling ultrasonic problems for the 2002 benchmark session," Review of Progress in Quantitative Nondestructive Evaluation, D.O. Thompson and D.E. Chimenti, Eds., American Institute of Physics, Melville, N.Y. , 22B,1776-1783, 2003.

The 3-D scattering amplitude component of a 2-D scatterer of length L with the end effects neglected can be written as shown and the measurement model is in the form given.

Ultrasonic Measurement Models

Relationship between 2-D and 3-D scattering amplitudes

$$A_{2D}\left(\mathbf{e}_i^{\beta},\mathbf{e}_s^{\alpha},\omega\right)=\sqrt{\frac{2i\pi}{k}}\,\frac{A_{3D}\left(\mathbf{e}_i^{\beta},\mathbf{e}_s^{\alpha},\omega\right)}{L}$$

which gives the alternate form

$$V_R(\omega)=s(\omega)\left[\int_{length,L}\hat{V}_0^{(1)}\hat{V}_0^{(2)}dx_2\right]A_{2D}\left(\mathbf{e}_i^{\beta},\mathbf{e}_s^{\alpha},\omega\right)\left[\sqrt{\frac{8\pi i}{k_2}}\,\frac{\rho_2 c_2}{\rho_1 c_1 S_A}\right]$$

Many model studies of cylindrical scatterers treat the scattering problem as a purely 2-D problem described by a 2-D scattering amplitude in the far field. We can relate such 2-D model results to our scattering problem, which is inherently 3-D in nature and write the measurement model in terms of a scalar 2-D scattering amplitude instead if we so desire.

Measurement Model for 2-D scatterers

$$V_R(\omega) = s(\omega)\left[\int_{length,L} \hat{V}_0^{(1)}\hat{V}_0^{(2)} dx_2\right]\frac{A_{3D}\left(\mathbf{e}_i^{\beta}, \mathbf{e}_s^{\alpha}, \omega\right)}{L}\left[\frac{4\pi}{-ik_2 S_A}\frac{\rho_2 c_2}{\rho_1 c_1}\right]$$

1) linear, reciprocal acoustic and elastic media (harmonic disturbances)

2) the output velocity for the transducer(s) can be written in a separable form

3) the pulser, receiver, transducers and cabling can be replaced by equivalent linear, time-shift invariant (LTI) systems.

4) the transducers are assumed to behave as piston radiators operating at high frequency

5) the incident waves are quasi-plane waves in the vicinity of the flaw

6) the scatterer is cylindrical flaw with incident and scattered wave directions in the plane of the cross-section, where the incident waves do not vary significantly in amplitude over the small cross-section of the flaw

Here is the **Thompson-Gray type of measurement model for 2-D scatterers** and the assumptions on which it is based.

Ultrasonic Measurement Models

General Measurement Model (Auld form)

$$V_R(\omega) = \frac{s(\omega)}{Z_r^A(\omega) v_A^{(1)}(\omega) v_B^{(2)}(\omega)} \int_{S_f} \left(\mathbf{t}^{(1)} \cdot \mathbf{v}^{(2)} - \mathbf{t}^{(2)} \cdot \mathbf{v}^{(1)} \right) dS$$

3-D small flaw (Thompson-Gray form)

$$V_R(\omega) = s(\omega) \left[\hat{V}_0^{(1)} \hat{V}_0^{(2)} \right] \left[A_{3D}\left(\mathbf{e}_i^{\beta}, \mathbf{e}_s^{\alpha}, \omega\right) \right] \left[\frac{4\pi}{-ik_2 S_A} \frac{\rho_2 c_2}{\rho_1 c_1} \right]$$

2-D small flaw (Schmerr-Sedov form)

$$V_R(\omega) = s(\omega) \left[\int_{length, L} \hat{V}_0^{(1)} \hat{V}_0^{(2)} dx_2 \right] \frac{A_{3D}\left(\mathbf{e}_i^{\beta}, \mathbf{e}_s^{\alpha}, \omega\right)}{L} \left[\frac{4\pi}{-ik_2 S_A} \frac{\rho_2 c_2}{\rho_1 c_1} \right]$$

In summary, we have three types of measurement models: (1) the Auld type of model that is applicable to a very wide range of problems and flaws, (2) a Thompson-Gray type of model suitable for the inspection of 3-D small flaws, and (3) A Thompson-Gray type of model developed by L. Schmerr and A. Sedov that is suitable for 2-D scatterers such as side-drilled holes.

Ultrasonic Measurement Models

Using Measurement Models

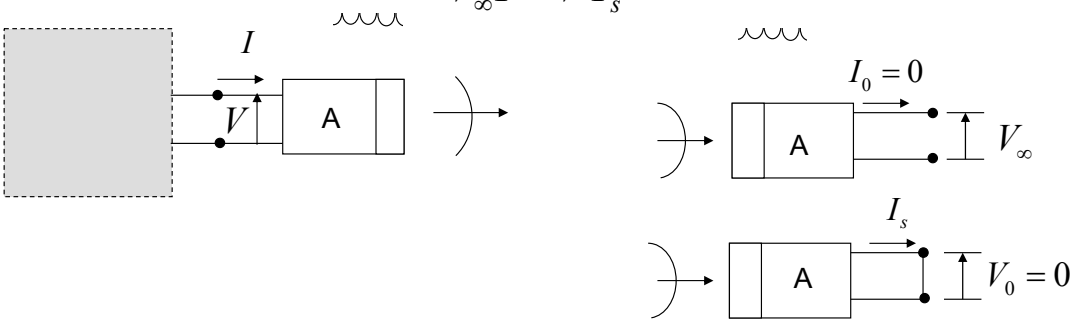
Is the system linear and reciprocal?

Electrical linearity checks of pulser/receiver

Cabling reciprocity checks

For a transducer, reciprocity implies the voltage and current measurements when it is being used as a sender and receiver must be related:

$$V_\infty I = V I_s$$

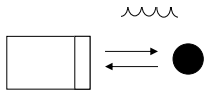


Linearity and reciprocity are two key assumptions in all the ultrasonic measurement models. We can, of course, do linearity checks on the pulser/receiver (and such checks are sometimes specified in inspection protocols). We have seen how to check reciprocity for our cable measurements. We can examine a transducer acting as a sender and receiver, as shown, and check its reciprocity with the relationship given above for the measured electrical signals. This relationship comes from writing the ratios V/I and I_s/V_{inf} in terms of the 2x2 transfer function components for the transducer on transmission and reception and assuming those components satisfy reciprocity. In that case these ratios are equal, leading to the above expression).

Ultrasonic Measurement Models

Is the transducer acting as a piston?

Compare responses with theoretical predictions in a reference experiment, e.g.



Is a quasi-plane wave assumption valid?

Compare $\int_{S_f} \left(\frac{\tau_{ij}^{(1)}}{v_A^{(1)}} \frac{v_i^{(2)}}{v_B^{(2)}} - \frac{v_i^{(1)}}{v_A^{(1)}} \frac{\tau_{ij}^{(2)}}{v_B^{(2)}} \right) n_j \, dS$ as computed with a non-paraxial beam model

with $\left[\frac{4\pi\rho_2 c_2}{-ik_2} \right] \int_{S_f} \hat{V}^{(1)} \hat{V}^{(2)} \mathbb{A}\left(\tilde{v}_j^{(1)}, \tilde{\tau}_{ij}^{(1)} \right) dS$

We can also check to see if a piston transducer model is appropriate by, for example, moving a small reflector along the axis of the transducer and comparing the on-axis fields with those predicted by a piston transducer model. For a Thompson-Gray type of model, which relies on the quasi-plane assumption, we could check the validity of that assumption by comparing the fields found in a general Auld type of model with the reduced forms appropriate for a quasi-plane wave (paraxial) model.

Ultrasonic Measurement Models

Is a small flaw assumption valid?

Compare $\int_{S_f} \hat{V}^{(1)}\hat{V}^{(2)} \mathbb{A}\left(\tilde{v}_j^{(1)}, \tilde{\tau}_{ij}^{(1)}\right) dS$

with $\left[\hat{V}_0^{(1)}\hat{V}_0^{(2)}\right] A_{3D}\left(\mathbf{e}_i^{\beta}, \mathbf{e}_s^{\alpha}, \omega\right)$ (3-D case)

or $\left[\frac{1}{L}\int_{length,L} \hat{V}_0^{(1)}\hat{V}_0^{(2)} dx_2\right] A_{3D}\left(\mathbf{e}_i^{\beta}, \mathbf{e}_s^{\alpha}, \omega\right)$ (cylindrical case)

The Thompson-Gray model also assumes the fields do not vary significantly over the flaw surface so that assumption can also be checked by a comparison with the more general underlying model terms.

Ultrasonic Measurement Models

What supporting measurements are needed?

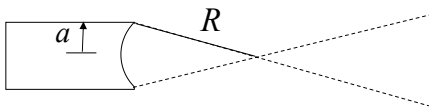
Velocity, Attenuation

Obtain in separate experiments

System factor

Obtain in separate experiment

Transducer effective radius, focal length



Obtain in separate experiments

An ultrasonic measurement model also relies on a number of supporting measurements. These include measurements of velocity and attenuation in the materials being inspected and the measurement of the system function (or its underlying components if we want to do a more in-depth characterization of the system as we have discussed). The transducer effective radius and focal length are also parameters that can be measured.

Ultrasonic Measurement Models

Summary

A measurement model form is available suitable for simulating most NDE inspections

To make these models useful one needs to have
- ultrasonic beam models
- flaw scattering models
- experimental determination of essential parameters

Once all these elements are available one has a very powerful tool for designing tests, optimizing components, and replacing expensive experimental procedures.

The measurement models we have discussed give us the capability to model many NDE experimental setups. These models can both be used as design tools and for understanding the signals seen in ultrasonic tests.

Measurement Model Examples

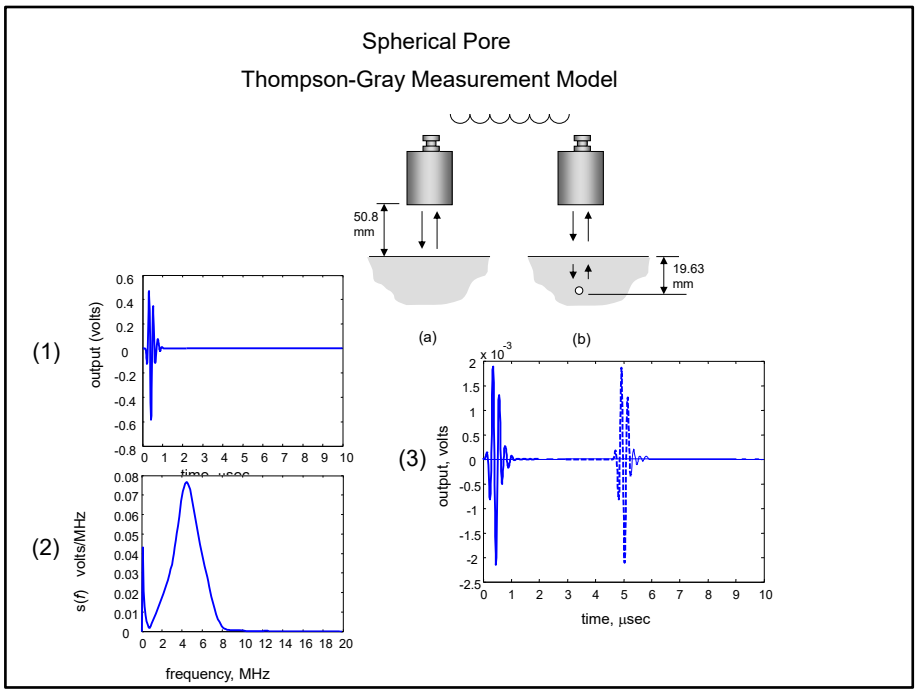
Here we will look at some comparisons of measurement model predictions with measured signals.

Learning Objectives

Comparison of the measured signals and the signals predicted by a measurement model for three common reference reflectors:

spherical pore
flat-bottom hole (flat crack)
side-drilled hole

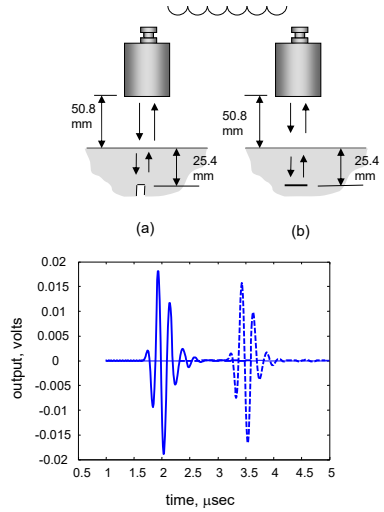
To see the quality of the signals predicted by our ultrasonic measurement models in comparison with actual experimental signals, we have shown some time domain responses for three types of commonly used reference reflectors.



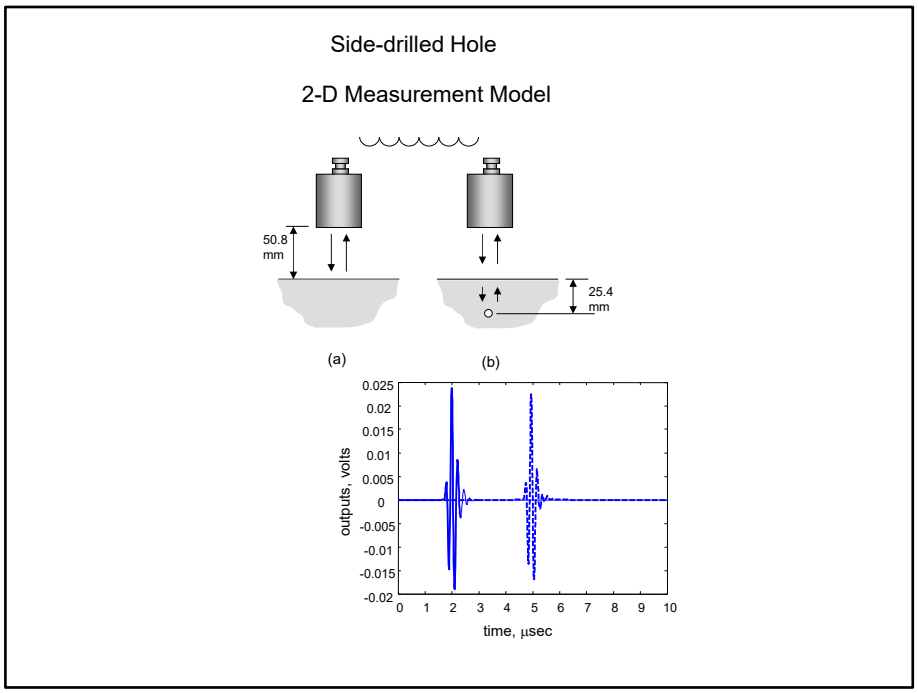
(1) The measurement model results for the pulse-echo response of a 0.6921 mm diameter spherical void, using the Thompson-Gray measurement model and the Kirchhoff approximation for the scattering amplitude. (2) The system function, as measured in setup (a), for the 12.7 mm diameter, 5 MHz transducer used. (3) A comparison of the measurement model signal (solid line) with the experimentally measured signal (dashed line).

Flat-bottom hole or flat crack at normal incidence

"Large" flaw measurement model



The measurement model results for the pulse-echo P-wave response of a 1.5875 mm radius flat-bottom hole (which also looks like a flat crack), where the system function was measured as in the previous case and the variation of the amplitude of the incident waves over the flat-surface was taken into account in the Kirchhoff approximation (Highly specular reflectors, like a flat-bottom hole or flat crack, are more sensitive to the small flaw assumption than are reflectors like the spherical pore just considered) . Shown is a comparison of the measurement model signal (solid line) with the experimentally measured signal (dashed line).



Comparison of the measurement model predictions for the pulse-echo P-wave response of a 1 mm diameter side-drilled hole with a 5MHz transducer (solid line) with the experimentally measured signal.

Special Topics – 1
Dispersion

Here we will briefly describe wave dispersion

Learning Objectives

Definition of dispersion

Group velocity

Examples

We will define dispersion and the group velocity of waves. We will also give an example of the effects of dispersion on waves as they propagate.

Dispersive Waves and the Group velocity

Recall that the wave number k, and the frequency, ω, were related to the (phase) velocity, c, through the relationship

$$k = \frac{\omega}{c}$$

ω … frequency in rad/sec

If the phase velocity, c, is a function of frequency then we say that the wave is dispersive and the wave number is

$$k(\omega) = \frac{\omega}{c(\omega)}$$

and we have

$$\frac{dk}{d\omega} = \frac{1}{c}\left[1 - \frac{\omega}{c}\frac{dc}{d\omega}\right] = \frac{1}{c_g}$$

where c_g is called the group velocity.

The group velocity can be shown to be the speed at which the energy in a dispersive wave propagates.

For propagation problems where the wave speed is a constant, the wave number varies linearly with the frequency. However, if the wave speed itself is a function of frequency, then the wave number is a more complex function of frequency and its derivative with respect to frequency is the reciprocal of the **group velocity**. Physically, the group velocity is the wave speed at which energy propagates in a wave. The frequency dependent wave speed itself is called the **phase velocity**.

Group velocity $$c_g = \frac{c}{1 - \frac{\omega}{c}\frac{dc}{d\omega}} = \frac{c}{1 - \frac{f}{c}\frac{dc}{df}}$$

where f is the frequency in cycles/sec (Hz)

If A(f) is the Fourier transform of a(t) then we have for a propagating non-dispersive plane wave traveling in the +x direction with a constant phase velocity, c,

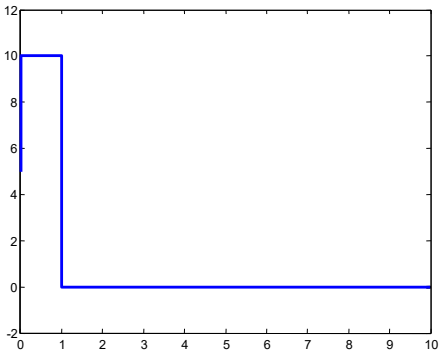
$$a(t - x/c) = \int_{-\infty}^{+\infty} A(f)\exp[2\pi i f x/c]\,df$$

However, if the phase velocity is a function of frequency then as the wave propagates with its energy traveling with the group velocity and its profile changes with increasing distance

$$a_d(t,x) = \int_{-\infty}^{+\infty} A(f)\exp[2\pi i f x/c(f)]\,df$$

Here is the group velocity expressed in terms of the phase velocity and its derivative. The superposition of non-dispersive harmonic waves through the use of the inverse Fourier transform produces a traveling wave in the time domain that always has the same profile. The superposition of dispersive harmonic waves, however, generates a traveling waveform whose shape changes.

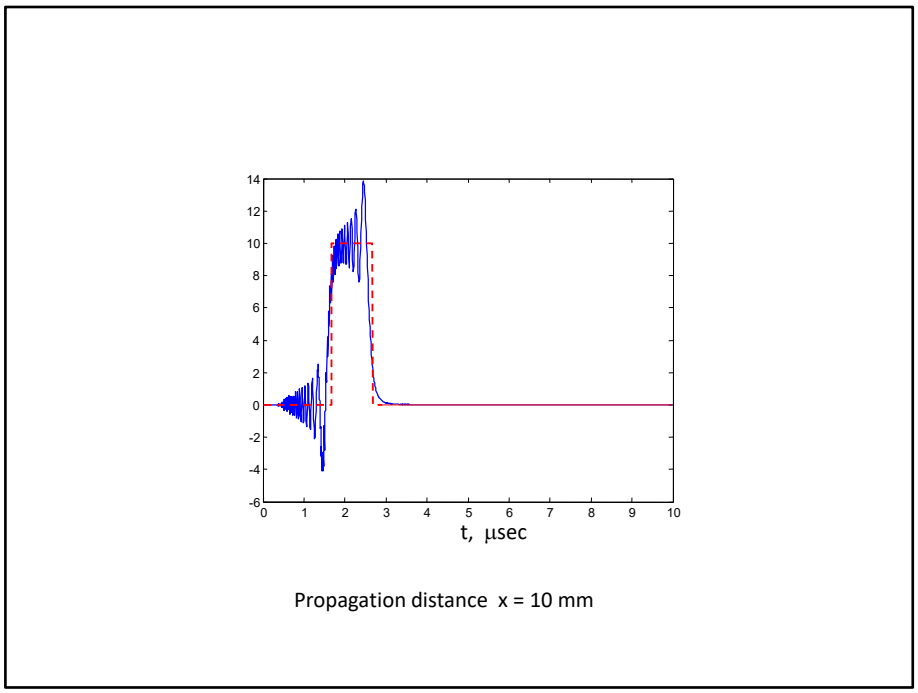
$c = 6\left[1 + f/50\right]$ mm/μsec f in MHz



t, μsec

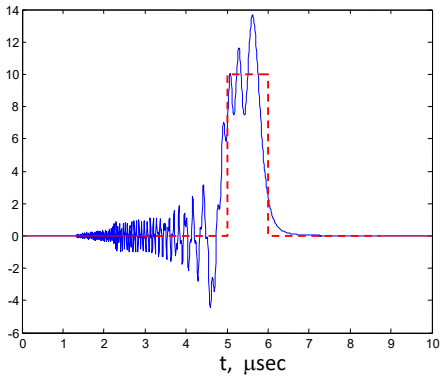
Propagation distance x = 0

To see an example of dispersion, consider a waveform which starts out at x = 0 with a “box” profile in time, as shown. Let the wave speed (of the medium it is propagating in) have the simple linear dependency on frequency, as shown.



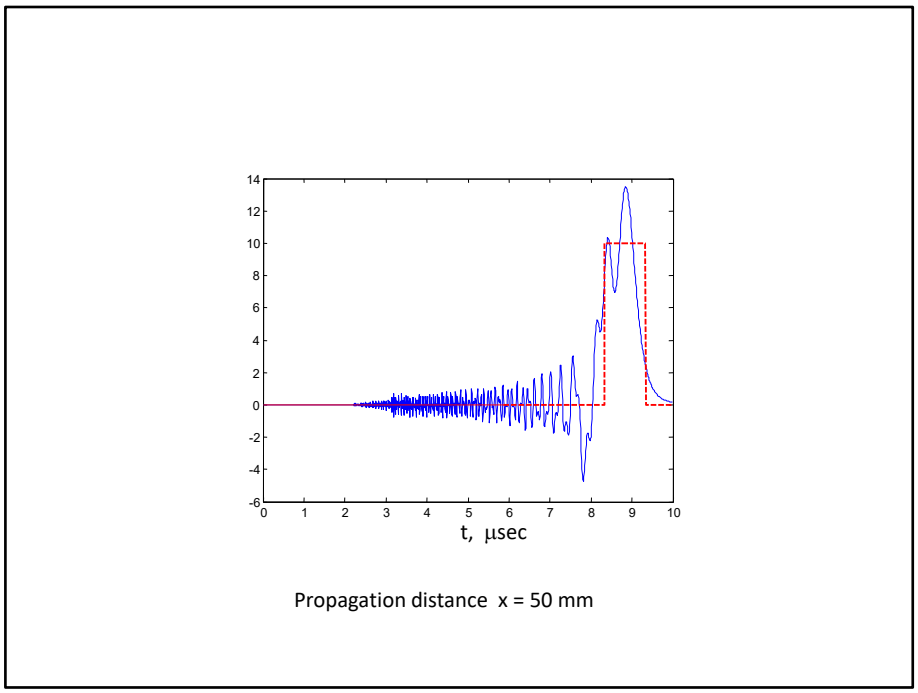
Propagation distance x = 10 mm

If one computes the inverse Fourier transform when the wave is at x = 10 mm, the waveform has changed dramatically, with significant ringing preceding and within the wave form itself. The dotted red line would be the waveform if it traveled in a non-dispersive medium at 6 mm/ microsec.



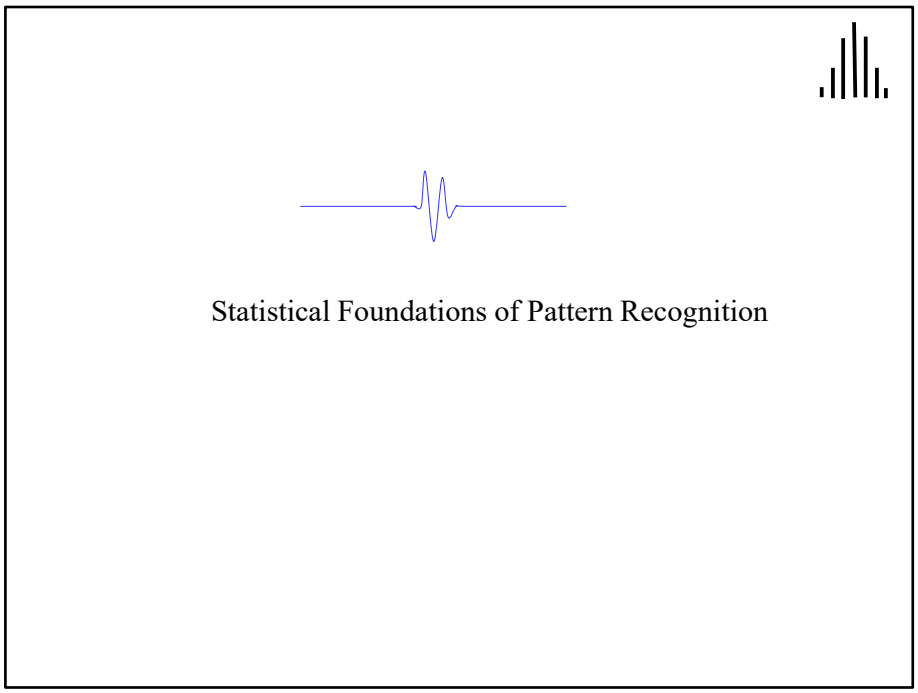
Propagation distance x = 30 mm

At x = 30 mm the wave distortions are even more pronounced.

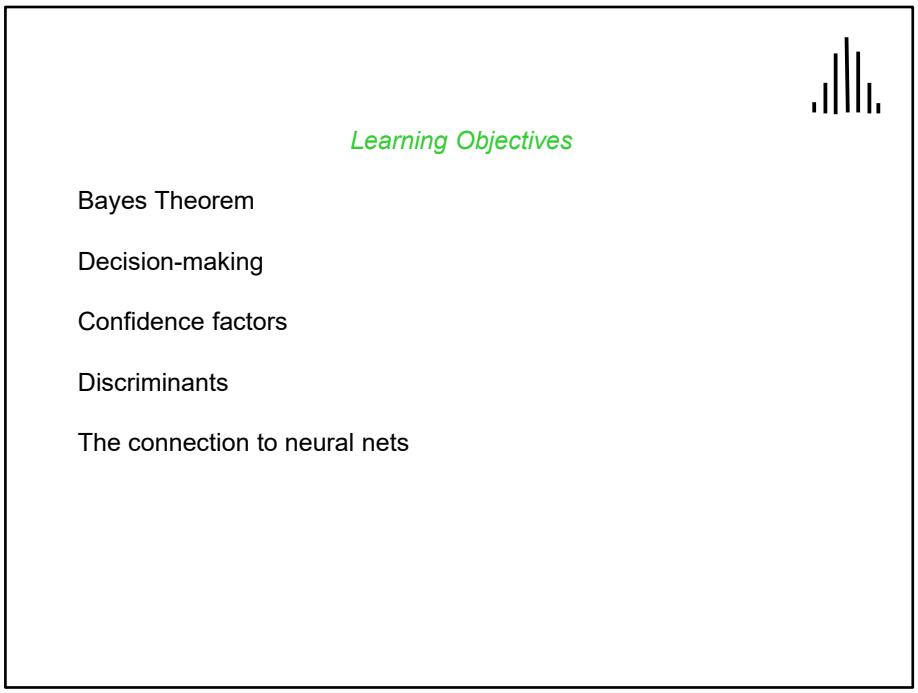


At x = 50 mm again there are continuing wave form changes. The large ringing see at earlier times comes from the fact that the higher frequencies travel with higher wave speeds in this case so they run ahead of the main wave form.

Waves traveling in plates, pipes, and shells (called **guided waves**) inherently are dispersive so that guided wave NDE inspections must deal with the effects of dispersion.

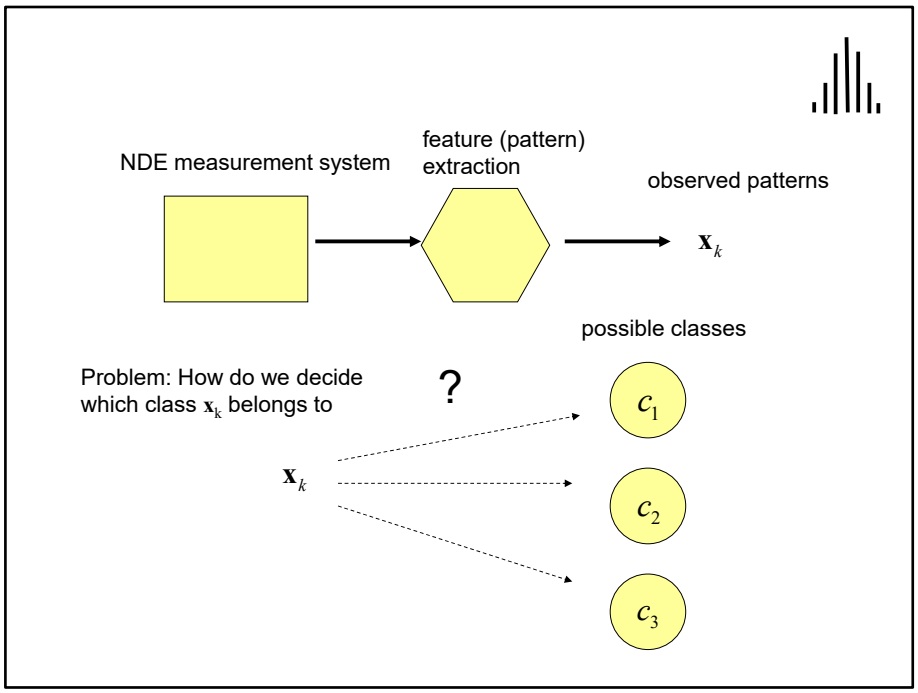


We will give a very brief overview here of how we can use statistical methods to evaluate the meaning of ultrasonic NDE flaw signals.



An ultrasonic NDE flaw measurement is an **indirect measurement** in that we normally have a measured output such as a voltage versus time trace received from the flaw which we must interpret to determine the flaw characteristics such as size, type (e.g. crack or inclusion) or material properties, etc. We can view many of these problems as pattern recognition problems. Statistical methods based on probabilities have long been used in this task so we will examine a statistical approach to the decision-making process in pattern recognition, based on **Bayes theorem**, which we will define shortly. We will go through some simple examples and introduce uncertainties in the decision-making process with the use of **confidence factors**.

We will see how the underlying probabilities can be replaced by quantities called **discriminants** and how we can learn such discriminants directly from the underlying data. We will connect this learning process to a simple form of a **neural network**. Neural nets are powerful pattern recognition tools that we will discuss more extensively in a separate presentation.



In an ultrasonic NDE flaw measurement system we can obtain some "raw" (i.e. unprocessed) flaw signals and then use those signals to extract some specific features from the signals in the form of observed patterns. The problem we want to consider here is: given an observed pattern of features and a set of possible classes of flaws from which it could come, how do we decide which class to choose? We will give a simple example shortly that will tie this general pattern recognition task more closely to an ultrasonic measurement but first we need to define some of the probabilities that we will be talking about when we use Bayes theorem as part of the decision-making process.

Bayesian (probabilistic) approach

Let $P(c_i)$ = apriori probability that a pattern belongs to class c_i, regardless of the identity of the pattern

$P(\mathbf{x}_k)$ = apriori probability that a pattern is $\mathbf{x}_k$, regardless of its class membership

$P(\mathbf{x}_k \mid c_i)$ = conditional probability that the pattern is $\mathbf{x}_k$, given that it belongs to class c_i

$P(c_i \mid \mathbf{x}_k)$ = conditional probability that the pattern's class membership is c_i , given that the pattern is $\mathbf{x}_k$

$P(\mathbf{x}_k, c_i)$ = the joint probability that the pattern is $\mathbf{x}_k$ and the class membership is c_i

A Bayes approach is based on probabilities, but there are number of different types of probabilities we must consider. There are two **a priori probabilities**. There is the a priori probability that the pattern we see comes from a specific class and the a priori probability that we will see a given pattern. There are also two **conditional probabilities**. There is the probability that a given pattern is present when a measurement is done on a particular class and the probability that a particular class is present given that we see a given pattern. Finally, there is a **joint probability** that both a given pattern and given class are both present.

In the next slide we will put some more specific flesh on the meaning of these probabilities.

Example: consider the case where there is one pattern value that is observed or not, and two classes (e.g. signal or noise)

x, c_1 *
~x, c_1
x, c_1 *
~x, c_2
x, c_2 #
x, c_1 *
x, c_2 #
~x, c_1
~x, c_1
x, c_1 *

(~x means x not observed)

Then

$P(x) = 6/10$
$P(c_1) = 7/10$
$P(c_2) = 3/10$
$P(x \mid c_1) = 4/7$
$P(x \mid c_2) = 2/3$
$P(c_1 \mid x) = 4/6$
$P(c_2 \mid x) = 2/6$
$P(c_1, x) = 4/10$ (see * s)
$P(c_2, x) = 2/10$ (see # s)

Consider the flaw detection problem of distinguishing between whether we are seeing a flaw signal or are simply seeing "noise." Here there are obviously two classes : c1: flaw signal and c2: noise (no flaw). Suppose we use the amplitude of the flaw voltage signal to make that decision and set a voltage threshold value above which we say we have observed a flaw, a pattern feature we will label as x, and below which we say we have not observed a flaw, a pattern feature we will label as ~x. Now, let us take ten measurements on samples containing either flaws or no flaws and suppose the ten results are as shown above. Then from examining those cases we can determine a priori probabilities. conditional probabilities, and joint probabilities, as shown. Of course these are very few samples so we should not take probability values based on such a small amount of data too seriously. They are simply meant to illustrate the meaning of the various probabilities we defined.

Bayes Theorem

$$P(c_i, \mathbf{x}_k) = P(\mathbf{x}_k \mid c_i) P(c_i)$$
$$= P(c_i \mid \mathbf{x}_k) P(\mathbf{x}_k)$$

or, equivalently, in an "updating form"

"new" probability of c_i , having seen $\mathbf{x}_k$ → $P(c_i \mid \mathbf{x}_k) = \frac{P(\mathbf{x}_k \mid c_i) P(c_i)}{P(\mathbf{x}_k)}$ ← "old" probability of c_i

Bayes theorem connects these probabilities since it says we can write the joint probability in terms of either sets of conditional probabilities and a priori probabilities, as shown. We can write this theorem in a form where given a set of a priori probabilities and a set of measurements, we can give an “updated” probability (from the original a priori probability that of a given class) that a given class is present, based on the conditional probability that we have a set of features from that class .

Bayes Theorem

$$P(c_i \mid \mathbf{x}_k) = \frac{P(\mathbf{x}_k \mid c_i) P(c_i)}{P(\mathbf{x}_k)}$$

Since $P(\mathbf{x}_k) = \sum_j P(\mathbf{x}_k \mid c_j) P(c_j)$

we can calculate $P(c_i \mid \mathbf{x}_k)$ if we know the probabilities

$P(c_j) \quad j = 1, 2, \ldots.$ and $P(\mathbf{x}_k \mid c_j) \quad j = 1, 2, \ldots$

In Bayes theorem it appears we need to have three probabilities to do an "updating" but in reality we only need the a priori probabilities of the classes present and conditional probabilities that a measured pattern feature comes from a particular class.

Now, consider our previous example

$P(x) = 6/10$
$P(c_1) = 7/10$
$P(c_2) = 3/10$
$P(x \mid c_1) = 4/7$
$P(x \mid c_2) = 2/3$
$P(c_1 \mid x) = 4/6$
$P(c_2 \mid x) = 2/6$
$P(c_1, x) = 4/10$
$P(c_2, x) = 2/10$

$$\begin{aligned} P(c_1, x) &= P(x \mid c_1) P(c_1) \\ &= (4/7)(7/10) = 4/10 \\ &= P(c_1 \mid x) P(x) \\ &= (4/6)(6/10) = 4/10 \end{aligned}$$

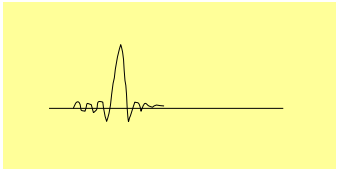
or, in the "updating" form

$$\begin{aligned} P(c_1 \mid x) &= \frac{P(x \mid c_1) P(c_1)}{P(x \mid c_1) P(c_1) + P(x \mid c_2) P(c_2)} \\ &= \frac{(4/7)(7/10)}{(4/7)(7/10) + (2/3)(3/10)} \\ &= \frac{(4/7)(7/10)}{(6/10) \longleftarrow P(x)} \\ &= 4/6 \end{aligned}$$

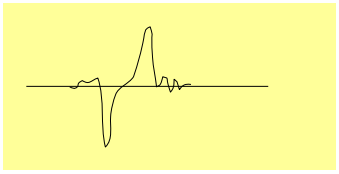
Here is our previous example of ten measurements, where we see that the various probabilities are indeed related through Bayes Theorem. Of course, this is an artificial "static" example where we have simply used Bayes theorem to relate the various probabilities from a set of known examples. The real power of Bayes theorem comes when we use the updating form of that theorem in a "dynamic" setting to provide an improved decision of what we are seeing based on new data as it comes in. We will give a specific example of this next.

As a simple example, consider trying to classify a flaw as a crack or a volumetric flaw based on the these two features:

x_1 : a positive leading edge pulse, PP



x_2 : flash points, FP



Cracks are typically more dangerous flaws than are volumetric flaws such as inclusions. Thus, suppose we consider trying to classify a flaw as a crack or non-crack (volumetric flaw). From our discussions of flaw scattering we know an isolated crack will produce a pair of negative and positive flashpoint signals in the time domain, as shown. A volumetric flaw, on the other hand, will often produce a large leading edge response, which may be positive or negative, depending on the acoustic impedance of the flaw relative to the surrounding material. A positive leading edge response will likely be easier to distinguish from a set of flashpoints, so let us choose a positive leading edge response and flashpoints as the two features we will use to classify an unknown flaw signal. Of course we will have to do some processing of the raw time domain signals to determine if either feature is present, but let us assume that processing is done and a criterion is chosen so that we can determine the existence of these features from the signals.

Assume:

P(crack) =0.5

P(volumetric) = 0.5

P(PP | crack) = 0.1 (cracks have leading edge signal that is always negative, so unless the leading edge signal is mistakenly identified, this case is unlikely)

P(PP | volumetric) = 0.5 (low impedance (relative to host) volumetric flaws have negative leading edge pulses and high impedance volumetric flaws have positive leading edge pulses, so assume both types of volumetric flaws equally likely)

P(FP | crack)= 0.8 (flashpoints is a features strongly characteristic of cracks, so make this probability high)

P(FP | volumetric) =0.05 (alternatively, make this a very low probability)

Now, we need to set up the probabilities present in Bayes Theorem. These could come with previous experience from tests done on samples with know flaws or they could come from educated estimates.

(1) Now, suppose a piece of data comes in and there is firm evidence that flashpoints (FP) exists in the measured response. Then what is the probability that the flaw is a crack?

$$P(crack \mid FP) = \frac{P(FP \mid crack)P(crack)}{P(FP \mid crack)P(crack) + P(FP \mid vol)P(vol)}$$
$$= \frac{(0.8)(0.5)}{(0.8)(0.5) + (0.05)(0.5)}$$
$$= 0.94118$$

Thus, we also have

$$P(vol \mid FP) = 0.05882$$

Now, let us use Bayes Theorem as we do testing on a sample with a flaw that is unknown. Here is an example where our processing of the signals gives a firm indication of flashpoints. Flashpoints are strong indicators of cracks so we see our updated estimate of the probability that the flaw is a crack is very high, certainly much higher than the 50/50 probability we started out with.

(2) Now, suppose another piece of data comes in with the firm evidence of a positive leading edge pulse (PP). What is the new probability that the flaw is a crack?

$$P(crack \mid PP) = \frac{P(PP \mid crack)P(crack)}{P(PP \mid crack)P(crack) + P(PP \mid vol)P(vol)}$$

$$= \frac{(0.1)(0.94118)}{(0.1)(0.94118) + (0.5)(0.05882)}$$

$$= 0.76191$$

and, hence $P(vol \mid PP) = 0.23809$

Note how the previous $P(crack \mid PP)$ was now taken as the new, apriori $P(crack)$ in this Bayesian updating

Now we use another data sample such as a measurement of the flaw from a different angle, where there is firm evidence of a positive leading edge signal. We see how the probability that the flaw is a crack has been reduced by this new data.

(3) Finally, suppose another data set comes in with firm evidence that the flashpoints (FP) do not exist (written as ~FP). What is the probability now that the flaw is a crack?

$$P(crack \mid \sim FP) = \frac{P(\sim FP \mid crack)P(crack)}{P(\sim FP \mid crack)P(crack) + P(\sim FP \mid vol)P(vol)}$$

$$= \frac{(0.2)(0.762)}{(0.2)(0.762) + (0.95)(0.238)}$$

$$= 0.403$$

and now

$$P(vol \mid \sim FP) = 0.597$$

Note: $P(FP \mid crack) = 0.8 \rightarrow P(\sim FP \mid crack) = 0.2$

$P(FP \mid vol) = 0.05 \rightarrow P(\sim FP \mid vol) = 0.95$

We could also have a data sample where there is firm evidence that flashpoints do not exist. We can again use Bayes theorem in this case. Of course the new estimate of the probability that the flaw is a crack, given that we are certain that flashpoints are not present in this case, has been significantly reduced.

In all three cases we must make some decision on whether the flaw is a crack or not. One possible choice is to simply look at the probabilities and decide $\mathbf{x}_k$ belongs to class c = c_j if and only if

$$P(c_j \mid \mathbf{x}_k) > P(c_i \mid \mathbf{x}_k)$$

for all $i = 1, 2, ..., N \quad i \neq j$

Since in our present example we only have two classes, we only have the two conditional probabilities $P(crack \mid \mathbf{x}_k), P(vol \mid \mathbf{x}_k)$

and since just $P(vol \mid \mathbf{x}_k) = 1 - P(crack \mid \mathbf{x}_k)$, our decision rule is

$$P(crack \mid \mathbf{x}_k) > 1 - P(crack \mid \mathbf{x}_k)$$

or $$P(crack \mid \mathbf{x}_k) > 0.5$$

Bayes theorem gives us updated probabilities but we still need to use those probabilities to make a decision on whether the flaw is a crack or not. We could simply choose the conditional probability output from Bayes theorem which is the largest. Since there are only two classes in this case, this means we would make a decision when the output conditional probability is bigger than 0.5.

Using this simple probability decision rule, in the previous three cases we would find

		decision
(1)	$P(crack \mid FP) = 0.941$	crack
(2)	$P(crack \mid PP) = 0.761$	crack
(3)	$P(crack \mid \sim FP) = 0.401$	volumetric

There is no reason, however, that we need to make a decision on the conditional probabilities $P(c_i \mid \mathbf{x}_k)$ by themselves.

We could synthesize a decision functions $g_i(\mathbf{x}_k)$ from such conditional probabilities and use the g_i instead. This is the idea behind what is called Bayes Decision Rule:

Based on this type of decision process, here are the classification decisions we would make after our three tests. However, there is no reason to only use the probabilities themselves. We could instead define a decision function which is based on those probabilities, leading to what is called **Bayes decision rule**.

Bayes Decision Rule

Decide $\mathbf{x}_k$ belongs to class $c = c_j$ if and only if

$$g_j(\mathbf{x}_k) > g_i(\mathbf{x}_k)$$

for all $i = 1, 2, ..., N \quad i \neq j$ where $g_i(\mathbf{x}_k)$

is the decision function for class c_i

If we have generated such a decision function here is Bayes decision rule.

Example: Suppose that not all decision errors are equally important. We could weight these decisions by defining the loss, l_{ij} that is sustained when we decide class membership is c_i when it is in reality class c_j . Then in terms of these losses we could also define the risk that $\mathbf{x}_k$ belongs to to class c_i as

$$R_i(\mathbf{x}_k) = l_{ii}P(c_i \mid \mathbf{x}_k) + \sum_{j \neq i} l_{ij}P(c_j \mid \mathbf{x}_k)$$

For our two class problem we would have

$$R_1(\mathbf{x}_k) = l_{11}P(c_1 \mid \mathbf{x}_k) + l_{12}P(c_2 \mid \mathbf{x}_k)$$
$$R_2(\mathbf{x}_k) = l_{21}P(c_1 \mid \mathbf{x}_k) + l_{22}P(c_2 \mid \mathbf{x}_k)$$

The decision rule in this case would be to decide $\mathbf{x}_k$ belongs to class c_1 if and only if

$$R_1(\mathbf{x}_k) < R_2(\mathbf{x}_k)$$

or, equivalently $\quad (l_{11} - l_{21})P(c_1 \mid \mathbf{x}_k) < (l_{22} - l_{12})P(c_2 \mid \mathbf{x}_k)$

Here is an example of setting up such a decision function by including estimates of the losses sustained when we do a misclassification to construct a risk function that we will use in our decision-making instead.

In the special case where there is no loss when we guess correctly, then $l_{11} = l_{22} = 0$. If, also it is equally costly to guess either c_1 or c_2 then $l_{12} = l_{21}$ and the decision rule becomes

$$-l_{21}P(c_1 \mid \mathbf{x}_k) < -l_{21}P(c_2 \mid \mathbf{x}_k)$$

or $$P(c_1 \mid \mathbf{x}_k) > P(c_2 \mid \mathbf{x}_k)$$

which is the simple decision rule based on conditional probabilities we discussed previously

Note that this risk function approach also includes the special case of just using conditional probabilities as a special case.

Now, consider our previous example again and let c_1 = crack, c_2 = volumetric flaw and suppose we choose the following loss factors:

$l_{11} = -1$ (a gain. If we guess cracks, which are dangerous, correctly we should reward this decision)

$l_{12} = 1$ (if we guess that the flaw is a crack and it is really volumetric, then there is a cost (loss) since we may do unnecessary repairs or removal from service)

$l_{21} = 10$ (if we guess the flaw is volumetric and it is really a crack, there may be a significant loss because of a loss of safety due to misclassification)

$l_{22} = 0$ (if we guess it is volumetric and it is , there might be no loss or gain)

To construct the risk function we need to choose the loss factors. These factors may come, for example, from historical data or from safety concerns. Here is a set of choices for the loss factors.

In this case we find the decision rule is – decide that a crack is present if

$$(-11.0)P(c_1 \mid \mathbf{x}_k) < (-1.0)P(c_2 \mid \mathbf{x}_k)$$

or $$\frac{P(c_1 \mid \mathbf{x}_k)}{P(c_2 \mid \mathbf{x}_k)} > 0.091$$

For our example then we have

		decision
(1)	$\frac{P(c_1 \mid \mathbf{x}_2)}{P(c_2 \mid \mathbf{x}_2)} = \frac{0.941}{0.059} = 15.9 > 0.091$	c_1 (crack)
(2)	$\frac{P(c_1 \mid \mathbf{x}_1)}{P(c_2 \mid \mathbf{x}_1)} = \frac{0.761}{0.239} = 3.18 > 0.091$	c_1 (crack)
(3)	$\frac{P(c_1 \mid \sim \mathbf{x}_2)}{P(c_2 \mid \sim \mathbf{x}_2)} = \frac{0401}{0.599} = 0.669 > 0.091$	c_1 (crack)

If we use this risk function in our decision-making process with the previous three test examples, here are the decisions. Note how the losses have significantly affected our final decision.

Bayes Theorem (Odds)

We can also write Bayes Theorem in terms of odds rather than probabilities by noting that for any probability P (condition, joint, etc.) we have the corresponding odds, O, given by

$$O = \frac{P}{1-P} \qquad (\text{or } P = \frac{O}{1+O})$$

Example: $O(c_i \mid \mathbf{x}_k) = \frac{P(c_i \mid \mathbf{x}_k)}{1-P(c_i \mid \mathbf{x}_k)}$

Using this definition of odds, Bayes Theorem becomes

$$O(c_i \mid \mathbf{x}_k) = LR \;\; O(c_i)$$

where $LR = \frac{P(\mathbf{x}_k \mid c_i)}{P(\mathbf{x}_k \mid \sim c_i)}$ is called the likelihood ratio

Instead of using probabilities we could also write Bayes theorem in terms of odds, since we are accustomed to dealing with odds in a number of venues. In this form we see that the odds are simply updated by the **likelihood ratio**, LR, which is defined here

Going back to our example with x_1 =PP, x_2 =FP

$$P(c_1) = 0.5 \qquad O(c_1) = \frac{0.5}{1-0.5} = 1$$

$$P(c_2) = 0.5 \qquad O(c_2) = \frac{0.5}{1-0.5} = 1$$

$$P(x_1 \mid c_1) = 0.1 \qquad O(x_1 \mid c_1) = \frac{0.1}{1-0.1} = 0.11111$$

$$P(x_1 \mid c_2) = 0.5 \qquad O(x_1 \mid c_2) = \frac{0.5}{1-0.5} = 1$$

$$P(x_2 \mid c_1) = 0.8 \qquad O(x_1 \mid c_1) = \frac{0.8}{1-0.8} = 4$$

$$P(x_2 \mid c_2) = 0.05 \qquad O(x_2 \mid c_2) = \frac{0.05}{1-0.05} = 0.0526$$

Then for our three cases:

Here is our previous example of a two class problem with the two features of flash points and positive leading edge pulse, where we show all the odds corresponding to our original probabilities.

(1) $O(crack \mid FP) = \frac{P(FP \mid crack)}{P(FP \mid \sim crack)} O(crack)$

$= \frac{0.8}{0.05}(1) = 16$ and $P(crack \mid FP) = \frac{16}{1+16} = 0.941$

(2) $O(crack \mid PP) = \frac{P(PP \mid crack)}{P(PP \mid \sim crack)} O(crack)$

$= \frac{0.1}{0.5}(16) = 3.2$ and $P(crack \mid PP) = \frac{3.2}{1+3.2} = 0.762$

(3) $O(crack \mid \sim FP) = \frac{P(\sim FP \mid crack)}{P(\sim FP \mid \sim crack)} O(crack)$

$= \frac{0.2}{0.95}(3.2) = 0.674$ and $P(crack \mid \sim FP) = \frac{0.674}{1+0.674} = 0.403$

Here are the results of our three experiments in terms of the resulting conditional odds.

As we see from this result we can update the probabilities according to Bayes Theorem by

$$O(c_i \mid \mathbf{x}_k) = \frac{P(\mathbf{x}_k \mid c_i)}{P(\mathbf{x}_k \mid \sim c_i)} O(c_i)$$

if the feature pattern $\mathbf{x}_k$ is observed and by

$$O(c_i \mid \sim \mathbf{x}_k) = \frac{P(\sim \mathbf{x}_k \mid c_i)}{P(\sim \mathbf{x}_k \mid \sim c_i)} O(c_i)$$

if the feature pattern $\mathbf{x}_k$ is not observed. We can combine these two cases as

$$O(c_i \mid \hat{\mathbf{x}}_k) = LR(\hat{\mathbf{x}}_k, c_i) O(c_i)$$

where

$$LR(\hat{\mathbf{x}}_k, c_i) = \frac{P(\hat{\mathbf{x}}_k \mid c_i)}{P(\hat{\mathbf{x}}_k \mid \sim c_i)}$$

and

$$\hat{\mathbf{x}}_k = \begin{cases} \mathbf{x}_k & \text{if } \mathbf{x}_k \text{ is observed} \\ \sim \mathbf{x}_k & \text{if } \mathbf{x}_k \text{ is not observed} \end{cases}$$

Here we show that we can use the same form of Bayes theorem if a feature is observed or not observed by combining the two cases into the same form but with different likelihood ratios.

Criticisms of this Probabilistic Approach

1. It does not include uncertainty in the evidence of the existence (or not) of the feature patterns
2. It is difficult to assign the apriori probabilities

To solve the first problem we will show how to introduce uncertainty with confidence factors

To solve the second problems, we will discuss the alternative use of discriminants

Notice that there are two problems with this simple use of Bayes theorem. First, we always assumed there was firm evidence (by which we meant certainty) that a feature was present or not. Second, we had to assign initially the a priori probabilities, but gave no way to actually do that assignment in a given problem. Thus, we will now try to address those problems.

Confidence Factors

Consider Bayes Theorem in the odds form

$$O(c_i \mid \hat{\mathbf{x}}_k) = LR(\hat{\mathbf{x}}_k, c_i) O(c_i)$$

In updating the odds, the likelihood ratio is based on being able to have firm evidence of the existence of the pattern $\mathbf{x}_k$ or not

$$LR(\hat{\mathbf{x}}_k, c_i) = \frac{P(\hat{\mathbf{x}}_k \mid c_i)}{P(\hat{\mathbf{x}}_k \mid \sim c_i)}$$

We can introduce uncertainty into this updating by letting a user (or a program) give a response R in the range [-1,1], where

R = 1 corresponds to complete certainty $\mathbf{x}_k$ is present
R = 0 corresponds to complete uncertainty that $\mathbf{x}_k$ is or is not present
R = -1 corresponds to complete certainty that $\mathbf{x}_k$ is not present

We can introduce uncertainty in the updating of odds with Bayes theorem by intruding a confidence factor, R, which varies from -1 to 1 as indicated here.

Then in updating the odds, we can replace the likelihood ratio, LR, by a function of LR and R that incorporates this uncertainty

$$O(c_i \mid \hat{\mathbf{x}}_k) = f(LR, R)\,O(c_i)$$

There are, however, some properties that this function f should satisfy. They are:

1. if R = 1 $\quad f = LR(\mathbf{x}_k, c_i)$

(if we are certain in the evidence of x_k , we should reduce to ordinary Bayes)

2. If R = -1 $\quad f = LR(\sim \mathbf{x}_k, c_i)$

(if we are certain x_k does not exist, again reduce to ordinary Bayes)

3. If LR = 0 , f = 0

(if the likelihood is zero, regardless of the uncertainty, R, the updated odds should be zero)

We then can replace the likelihood ratio in Bayes theorem with a modified likelihood function that includes this confidence factor. Here are some properties such a function should have.

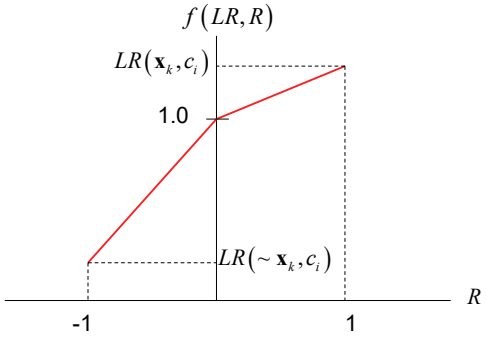
A popular choice that appears in the literature is to choose

$$f(LR,R) = LR(\hat{\mathbf{x}}_k, c_i)|R| + (1-|R|)$$

where, if R $\in$ [0,1] $\quad LR(\hat{\mathbf{x}}_k, c_i) = LR(\mathbf{x}_k, c_i)$

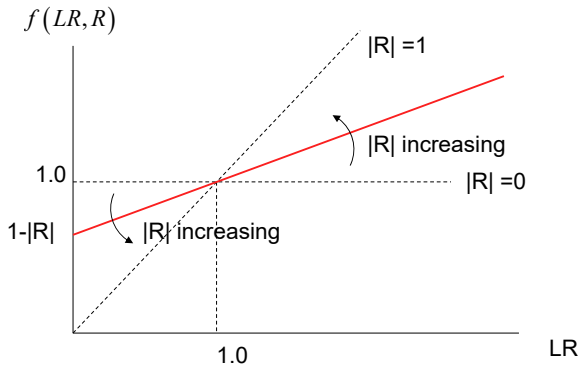
and, where, if R $\in$ [-1,0] $\quad LR(\hat{\mathbf{x}}_k, c_i) = LR(\sim \mathbf{x}_k, c_i)$

If we plot this function, we see the effects of R



Here is a piece-wise linear function one often sees in the literature.

Although this is a simple function to use, there is a problem with it which we can see if we plot f versus LR for different |R|. (Note that LR $\in [0, \infty)$)

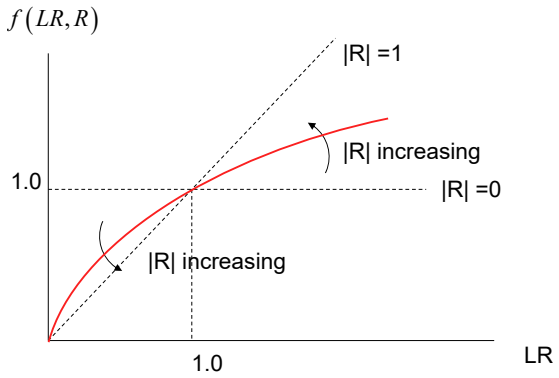


At LR = 0 the function f does not go to zero as we said it should (see property 3 discussed above). To remedy that problem, we need to choose a nonlinear function.

However, this choice does not satisfy the third property we gave earlier. Thus, we need to choose a different, nonlinear function.

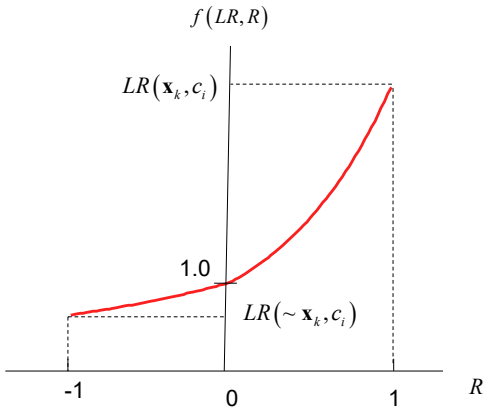
One choice that satisfies all three properties f should have is

$$f(LR, R) = (LR)^{|R|}$$



Here is one choice of such a nonlinear function which does have all the three properties listed earlier.

This gives a dependency on R that is nonlinear



This function is also nonlinear in R

With this choice of f, we would have

$$O(c_i \mid \hat{\mathbf{x}}_k) = LR^{|R|} O(c_i)$$

However, if one wants to work in terms of probabilities, not odds, we have

$$P(c_i \mid \hat{\mathbf{x}}_k) = \frac{P(c_i)}{P(c_i) + (LR)^{-|R|}(1 - P(c_i))}$$

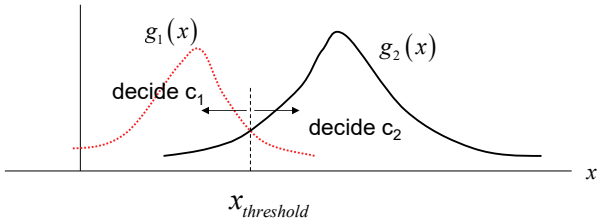
with $$LR = \frac{P(\hat{\mathbf{x}}_k \mid c_i)}{P(\hat{\mathbf{x}}_k \mid \sim c_i)}$$

and $$\hat{\mathbf{x}}_k = \begin{cases} \mathbf{x}_k & R > 0 \\ \sim \mathbf{x}_k & R < 0 \end{cases}$$

With this choice of the function then we can write Bayes theorem either in terms of odds or probabilities in forms that allow us to adjust the updating in a manner which allows us to incorporate uncertainty that the features are present in the decision-making process.

Bayes theorem, even in this modified form to take into account uncertainty in the evidence, still requires us to have apriori probability estimates and those may be difficult to come by. How do we get around this?

Consider our two class problem where we have classes (c_1, c_2) and where $\mathbf{x}_k$ =x is a single feature (pattern). According to Bayes decision rule we could decide on c_1 (or c_2) if $g_1(x) > g_2(x)$ (or $g_2(x) > g_1(x)$). For example, suppose both g_1 and g_2 were unimodal, smooth distributions. Then we might have:



Then we see the decision rule is really just

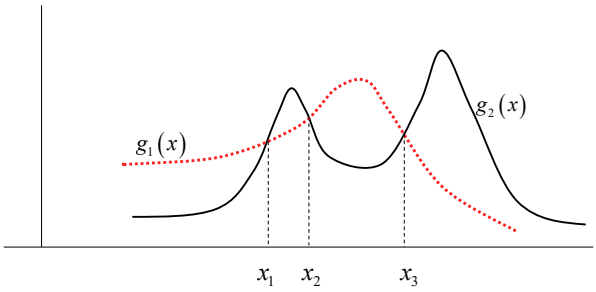
$$class = c_1 \quad if \; x < x_{threshold}$$
$$class = c_2 \quad if \; x > x_{threshold}$$

Now, let us turn our attention to the problem of having to give the a priori probabilities. In the two class problem we have been examining where we use Bayes decision rule, consider if there was only one feature, x, being used. Then the decision functions might look like two smooth functions having peaks at different values of x so that we could distinguish between the two classes. However, if instead of using the functions themselves to make a decision, we see we could instead just use a threshold value for x (where the two functions are equal) to make the decision.

Thus, if we had a way of finding $x_{threshold}$, which serves as a *discriminant*, we could make our decisions and not have to even consider the underlying probabilities!

However, we have not really eliminated the probabilities entirely since they ultimately determine the errors made in the decision making process.

Note that in the more general multi-modal decision function case, several discriminants may be needed:



If we had a way to determine the threshold in this single feature case, we could then base our decisions directly on that threshold value so that value serves as a **discriminant**. In cases where the behavior of the underlying decision functions were more complex, we might need several discriminants as shown in this example.

If we take the $g_i(x)$ to be just the probability distributions $P(c_i \mid x)$

then recall that Bayes decision rule says that $\mathbf{x}_k$ belongs to class $c = c_i$ if and only if

$$P(c_i \mid x) > P(c_j \mid x)$$

for all $j = 1, 2, ..., N \quad j \neq i$

or, equivalently $P(c_i \mid x)P(x) > P(c_j \mid x)P(x)$

which says that $P(c_i, x) > P(c_j, x)$

so that also $P(x \mid c_i)P(c_i) > P(x \mid c_j)P(c_j)$

If x is a continuous variable, then we can associate probability distributions with quantities such as $p(c_i, x)$ and $p(x \mid c_i)$ and so we expect that the discriminants are dependent on the nature of these distributions. We will now examine closer that relationship.

Here is Bayes decision rule for one feature x and where we have multiple possible classes, written in terms of probabilities. If x is a continuous variable than we could replace these discrete probabilities with probability distributions (as a function of x) and try to determine the discriminants from these distributions.

Probability Distributions and Discriminants

First, consider the 1-D case where $\boldsymbol{x} = x$ and where we assume the distributions are Gaussians, i.e.

$$p(x, c_i) = \left(\sigma_i \sqrt{2\pi}\right)^{-1} \exp\left[-(x-\mu_i)^2 / 2\sigma_i^2\right] P(c_i)$$

where

$\mu_i =$ mean value of x for class c_i

$\sigma_i =$ standard deviation for class c_i

If we assume $P(c_i) = P(c_j),\ \sigma_i = \sigma_j = \sigma$ then Bayes decision rule says that x belongs to class c_i if and only if

$$\frac{\exp\left[-(x-\mu_i)^2 / 2\sigma\right]}{\exp\left[-(x-\mu_j)^2 / 2\sigma\right]} > 1$$

Now, consider the case where there is one feature, x, which has continuous values, and assume the joint probability distributions for the various cases are Gaussians with different mean values and standard deviations. Note that Gaussians are smooth, peaked functions of the type we considered previously. Then if, for example, we assume the a priori probabilities are the same for all the classes then we can write the Bayes decision rule solely in terms of the properties of these Gaussians.

or, equivalently, x belongs to class c_i if and only if

$$(x-\mu_i)^2 < (x-\mu_j)^2$$

for all $j = 1, 2, ..., N \quad j \neq i$

This is just the basis for the **nearest cluster center** classification method

However, it is only the mean values of the Gaussians that determine the class, so those mean values act as discriminants and we could write our decision process in a form which is called the **nearest center cluster classification method**. Thus, here is a simple example where we can replace the underlying probabilities with dicriminants.

Now, consider the more general case of N-dimensional features but keep the assumption of Gaussian distributions. Then

$$p(\mathbf{x}, c_i) = \left[(2\pi)^{N/2} |\mathbf{\Sigma}_i|^{1/2} \right]^{-1} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \mathbf{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right] P(c_i)$$

where

$\boldsymbol{\mu}_i$ … N-component mean vector for class c_i

$\mathbf{\Sigma}_i$ … N x N covariance matrix

Let us generalize this problem by considering the more realistic case of multiple features and multiple classes. However, we will continue to use Gaussians which now must be written in terms of vectors and matrices. Do not be put off by these more complex forms since we will see much of this complexity can reduce eventually to a more understandable result.

Bayes decision theory says that **x** belongs to class c_i if and only if

$$\frac{P(c_i)\left[(2\pi)^{N/2}\left|\boldsymbol{\Sigma}_i\right|^{1/2}\right]^{-1}\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right]}{P(c_j)\left[(2\pi)^{N/2}\left|\boldsymbol{\Sigma}_j\right|^{1/2}\right]^{-1}\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)\right]} > 1$$

Now, suppose we are on the boundary between c_i and c_j and also suppose that

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_j = \sigma^2\mathbf{I}$$

where $\mathbf{I}$ is the unit matrix. Then

$$\frac{P(c_i)\exp\left[(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)/2\sigma^2\right]}{P(c_j)\exp\left[(\mathbf{x}-\boldsymbol{\mu}_j)^T(\mathbf{x}-\boldsymbol{\mu}_j)/2\sigma^2\right]} = 1$$

Taking the *ln* of this equation then we have

Here is Bayes decision theory in terms of these Gaussians. We will assume all the standard deviation matrices are identical in terms of a scalar standard deviation value and look at the case when we are on the decision boundaries between classes, since we know that is where a discriminant can exist between those classes. Taking the natural log of the expression seen in Bayes decision theory then gives the results seen on the next slide.

$$\ln\left[\frac{P(c_i)}{P(c_j)}\right] - (\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)/2\sigma^2 + (\mathbf{x}-\boldsymbol{\mu}_j)^T(\mathbf{x}-\boldsymbol{\mu}_j)/2\sigma^2 = 0$$

which can be expanded out to give

$$2\ln\left[\frac{P(c_i)}{P(c_j)}\right] + 2\mathbf{x}^T(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)/\sigma^2 + (\boldsymbol{\mu}_j^T\boldsymbol{\mu}_j - \boldsymbol{\mu}_i^T\boldsymbol{\mu}_i)/\sigma^2 = 0$$

However, these are just the equations of the hyperplanes

$$\mathbf{x}^T\mathbf{w}_{ij} = \mathbf{b}_{ij}$$

with

$$\mathbf{w}_{ij} = (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)/\sigma^2$$

$$\mathbf{b}_{ij} = (\boldsymbol{\mu}_i^T\boldsymbol{\mu}_i - \boldsymbol{\mu}_j^T\boldsymbol{\mu}_j)/2\sigma^2 - \ln\left[\frac{P(c_i)}{P(c_j)}\right]$$

The form of this result can be expanded out and written in matrix-vector form which shows that the boundaries defined between the classes are hyperplanes that are characterized by **w** and **b** matrices.

$$\mathbf{x}^T \mathbf{w}_{ij} = \mathbf{b}_{ij}$$

The $\mathbf{w}_{ij}$ and the $\mathbf{b}_{ij}$ here determine the hyperplanes separating the classes and hence are discriminants. If we can find a way to determine these discriminants directly, we need not deal with the underlying probabilities that define them. We will now examine ways in which we can find such hyperplanes (or hypersurfaces in a more general context).

The **w** and **b** matrices are our discriminants. Notice that the **b** matrices depend on the a priori probabilities but if we can find these discriminants directly then we do not need to know those a priori probabilities. Thus, we now need to determine a way to obtain such discriminants. we will give now a simple example of how to do that.

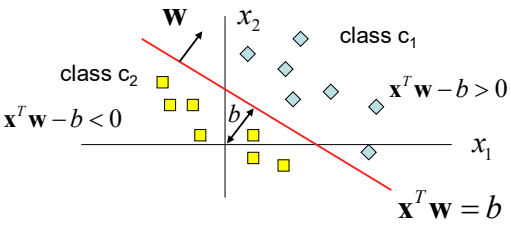
Learning Discriminants with Neural Networks

Suppose that we have a two class problem and a linear discriminant is able to distinguish between observed patterns, **x**, of either class. Such a problem is said to be *linearly separable*.

Geometrically, we have the situation where we can place a discriminant hyperplane

$$\mathbf{x}^T\mathbf{w} = b$$

between patterns, **x**, of either class. For example, for **x** = (x_1 , x_2)



= pattern observed from class c_2

= pattern observed from class c_1

In the following slides we will examine a two class problem where we use two features simultaneously to determine the class of a flaw. If we can define a line in this two dimensional feature space which acts as a discriminant between the two classes then we say that this is a **linearly separable classification problem**. The discriminating line here can be defined by a 2-D vector **w** = [w_1, w_2] which is normal to the line and a scalar b which represents the perpendicular distance from the origin to the line.

We will show that we can define a learning procedure that can directly determine the separating line discriminant and show how this procedure is similar to a neural network model where we use only a single model of a “neuron”. In a later presentation , we will generalize these ideas to more detailed neural networks consisting of many “neurons”

Learning to distinguish between these two classes then consists of finding the values of $\mathbf{w}$, b that will separate the observed patterns.

Note that we can augment the vector $\mathbf{x} = (x_1, x_2, ..., x_n)$

and the weight vector $\mathbf{w} = (w_1, w_2, ..., w_n)$

by redefining them as $\mathbf{w} \triangleq (w_1, w_2, ..., w_n, b)$

$\mathbf{x} \triangleq (x_1, x_2, ..., x_n, -1)$

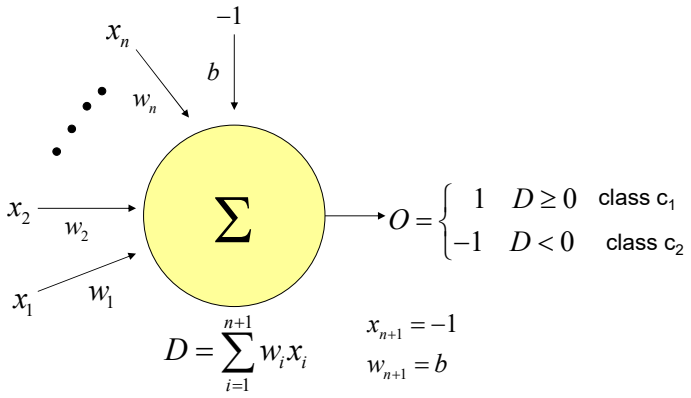
Then the equation of the hyperplane in terms of these augmented vectors becomes

$$\mathbf{x}^T \mathbf{w} = 0$$

To connect this linearly separable problem more closely to neural networks, let us incorporate the b constant into the weight vector, **w**, and similarly augment the vector **x** as shown. Then the hyperplane takes the simple form shown.

$$\mathbf{x}^T\mathbf{w} = 0$$

This equation can be related to neural network ideas since we can view the process of making a decision c_1 or c_2 as similar to the firing (or not firing) of a neuron based on the activity level of the inputs:



Now, the question is, how do we determine the unknown "weights", **w,** b ?

In this form we can relate our very simple discrimination problem to that of a “neuron.” Real neurons receive inputs from many other neurons and they “fire” when the input activity is high, thus exciting many other neurons. If we take our inputs as the elements of the augmented vector **x**, multiplied by the augmented weights in **w**, then we can sum those incoming weighted inputs into a factor, D, which represents the net activity coming into the neuron. If that factor is positive we can provide a high output of +1 (i.e. the “neuron fires”) or if the factor is negative we can provide a low output of -1 (i.e. the “neuron does not fire”). We can relate this process of firing or not to the existence of a flaw of class $c_1$or c_2. A neural network can be a large collection of interconnected “neurons” of this type (or other types). If we have a way to learn all the discriminants in such a network (called the **network weights**) then we have a powerful way to do pattern recognition on complex problems as well as perform other complex tasks.

We will talk more about neural nets later and describe the way we can learn to obtain the network weights but now let us describe a learning algorithm for our simple linearly separable classification problem.

Two Category Training Procedure

Given an extended weight vector $\mathbf{w} \triangleq (w_1, w_2, ..., w_n, b)$

and an extended feature vector $\mathbf{x} \triangleq (x_1, x_2, ..., x_n, -1)$

the following steps define a two class error correction algorithm

Definition: Let $\mathbf{w}_k$ be the weights associated with $\mathbf{x}_k$, " feature training vectors" for cases where the class of each $\mathbf{x}_k$ is known

(1) let $w_1 = (0, 0, ..., 0)$ (actually, w_1 can be arbitrary)

(2) Given $\mathbf{w}_k$, the following case rules apply

$case\,1: \quad \mathbf{x}_k \in c_1 \,(class\ c_1)$

$a.\ if\ \mathbf{x}_k \cdot \mathbf{w}_k \geq 0, \mathbf{w}_{k+1} = \mathbf{w}_k$

$b.\ if\ \mathbf{x}_k \cdot \mathbf{w}_k < 0, \mathbf{w}_{k+1} = \mathbf{w}_k + \lambda \mathbf{x}_k$

$case\,2: \quad \mathbf{x}_k \in c_2 \,(class\ c_2)$

$a.\ if\ \mathbf{x}_k \cdot \mathbf{w}_k < 0, \mathbf{w}_{k+1} = \mathbf{w}_k$

$b.\ if\ \mathbf{x}_k \cdot \mathbf{w}_k \geq 0, \mathbf{w}_{k+1} = \mathbf{w}_k - \lambda \mathbf{x}_k$ where $\lambda > 0$

We can define a two category (two class) learning procedure for a problem where the two classes can be separated in a feature space **x** by a hyperplane. Such a problem is called a **linearly separable classification problem**. We start with a given set of weights which can all be zero, for example, and a set of examples (NDE tests) where we have a set of features extracted from a known flaw type. We can call these examples the feature training vectors and define a learning procedure, as shown above, where we change the weights depending on the feature vector values and the type of class present. The procedure shown above is relatively simple where we either leave the weights unchanged or change them by either adding or subtracting the feature vector values multiplied by positive constant we choose. It can be shown this learning procedure must eventually stop and produce a set of weights which separates the two classes.

Two Category Learning Theorem

If c_1 and c_2 are linearly separable and the two class training procedure is used to define $\mathbf{w}_k$, then there exists an integer $t \geq 1$ such that $\mathbf{w}_t$ linearly separates c_1 and c_2 and hence $\mathbf{w}_{t+k} = \mathbf{w}_t$ for all positive k.

Ref: Hunt, E.B., **Artificial Intelligence**, Academic Press, 1975.

Remark: λ can be a constant or can take on particular forms. For example:

$$\lambda = \lambda_k = \frac{\alpha\, \mathbf{w}_k \cdot \mathbf{x}_k}{|\mathbf{x}_k|^2} \qquad (0 < \alpha < 2)$$

can be used and the algorithm still converges. This often speeds up the convergence

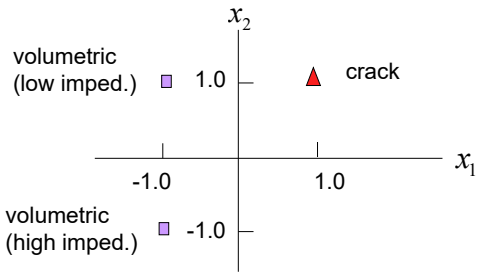
Ref: Nilsson, N, **Learning Machines**, McGraw Hill, 1965.

Here is a statement of the so-called **two category learning theorem**. As you can see, it has been known for some time. It was not until the 1980s that more powerful learning methods became readily available for neural networks, which allowed their use in problems much more complex and practical than the linearly separable case discussed here.

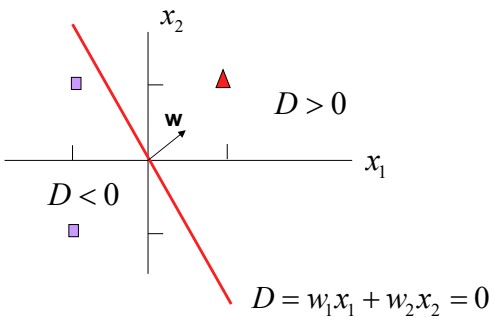
Example: Determine if an ultrasonic signal is from a crack or a volumetric flaw based on the following two features:

x_1 = " has flashpoints" 1 = "yes", -1 = "no"
x_2 = " has negative leading edge pulse" 1 = "yes", -1 = "no"

	crack	volumetric (low impedance)	volumetric (high impeance)
x_1	1	-1	-1
x_2	1	1	-1



Now, let's show a simple example of our two class flaw classification problem where we use as two features the existence of flashpoints and a negative leading edge pulses (we could also use the positive leading edge pulse feature considered previously, so this change is not important). We see in the feature space these distinguishing features can indeed be separated by a line (in fact, many lines) which we could obtain by inspection. However, we want to use our two category learning theorem to choose such a line, so we can see the learning process in action.



For simplicity, we will take b = 0, λ = 1 so the learning procedure is

1. Give an training example (x_1 , x_2)

2. If $D \geq 0$ ask "is it a crack" (Y or N)
   If $D < 0$ ask "is it volumetric" (Y or N)

3. If error (N) and $D \geq 0$ $\quad \mathbf{w}_{k+1} = \mathbf{w}_k - \mathbf{x}_k$
   If error (N) and $D < 0$ $\quad \mathbf{w}_{k+1} = \mathbf{w}_k + \mathbf{x}_k$

To make things very simple we will choose b=0 and λ= 0, but these choices are not necessary. Shown is the learning procedure for this case.

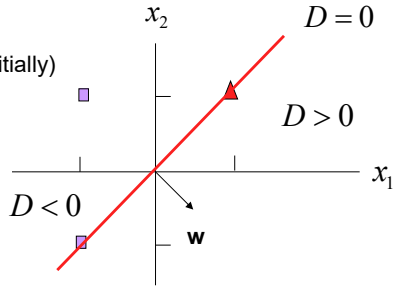
Suppose for this case we have the following training set:

1. $x_1 = -1$, $x_2 = 1$ (vol)
2. $x_1 = 1$, $x_2 = 1$ crack
3. $x_1 = -1$, $x_2 = -1$ (vol)
4. ….

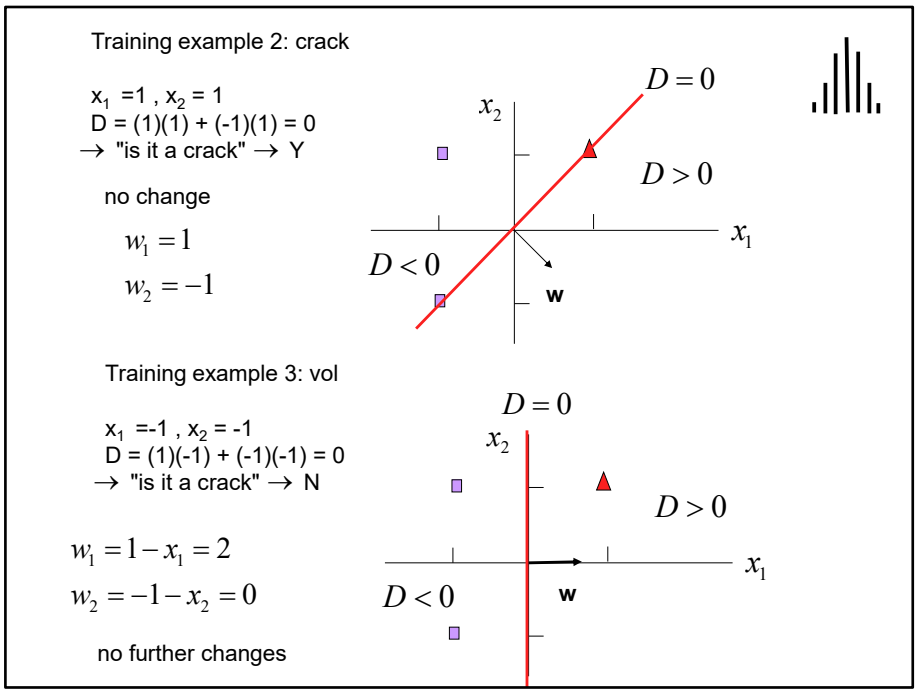
Training example 1: vol

x_1 =-1 , x_2 = 1
D = 0 (since $w_1 = w_2 = 0$ initially)
→ "is it a crack" → N

$$w_1 = 0 - (x_1) = 1$$
$$w_2 = 0 - (x_2) = -1$$



Now, suppose we have a set of examples that we use to learn a set of appropriate weights. Shown are the first three cases in this training set and how the weights are changed when we present the first training example to the learning procedure. The separating line that is generated is shown. Note that we have not yet separated the two cases because the high impedance volumetric flaw must give us $D<0$, not $D \leq 0$ as seen here.



Now, we present the second and third training examples. In the second example there is no change to the weights but in the third case the line now completely separates the two classes. Further training examples will not make any changes to this line, so we have learned the discriminants.

Note that this classifier can also handle situations other than which it is trained on. This "generalization" ability is a valuable property of neural nets.

For example, suppose we let

x_1 = 1 "definitely has flash points"
0.5 "probably has flash points"
0 "don't know"
-0.5 "probably does not have flash points"
-1 "definitely does not have flash point"

similarly for x_2

Now suppose we give our trained system an example it hasn't seen before such as a crack where

$x_1 = 0.5$ " probably has flash points"
$x_2 = 0.5$ " probably has a negative leading edge pulse"

$$D = (2)(0.5) + (0)(0.5) = 1 \geq 0$$

$\rightarrow$ " is it a crack" $\rightarrow$ Y (which is correct)

An important feature of our classifier is that it can handle situations other than just those it is trained on. For example, suppose we give the x-values something other than the +1 and -1 values used in training. In the example shown we see the classifier still works. This ability to generalize to cases not seen in the training samples is what makes neural networks such powerful tools.

References

Sklansky, J., and G. Wassel, *Pattern Classifiers and Trainable Machines*, Springer Verlag, 1981

Pao, Y.H., *Adaptive Pattern Recognition and Neural Networks*, Addison Wesley, 1989

Gale, W.A., Ed., *Artificial Intelligence and Statistics*, Addison Wesley, 1986

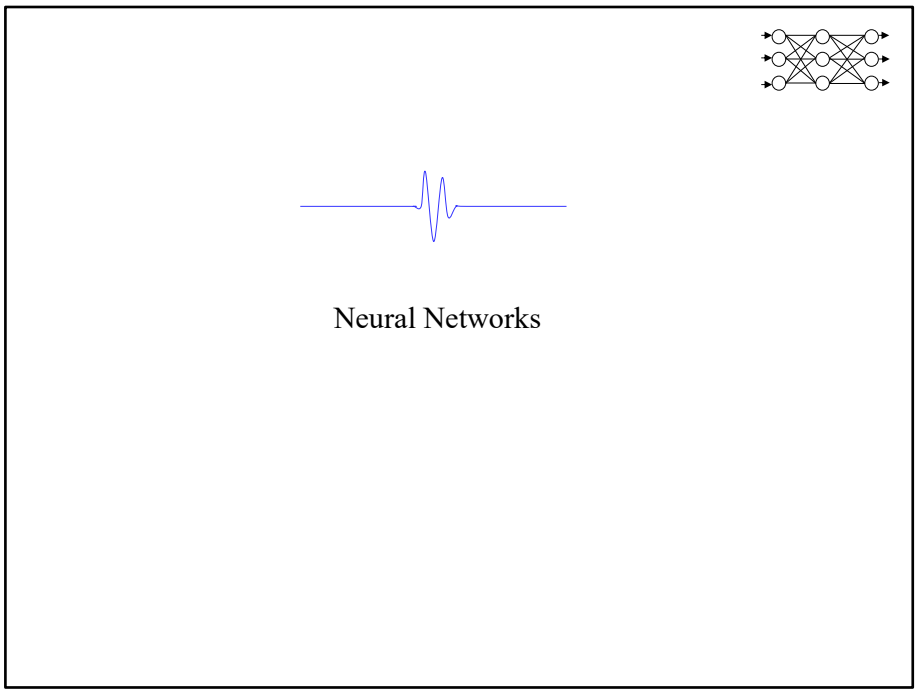
Duda, R.O, Hart, P.E., and D.G. Stork, *Pattern Classification*, 2nd Ed., John Wiley, 2001

Fukunaga, K. *Statistical Pattern Recognition*, Academic Press,1990.

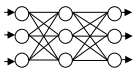
Webb, A, *Statistical Pattern Recognition*, 2nd Ed., John Wiley, 2002.

Nadler, M., and E.P. Smith, *Pattern Recognition Engineering*, John Wiley, 1993.

Here are some references on statistical pattern recognition. Later, we will give some references on neural networks.



Neural networks have become an important and powerful tool for analyzing signals so they are a valuable concept that can be used to evaluate the results of ultrasonic NDE inspections.



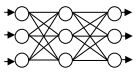
Learning Objectives

Characteristics of neural nets

Supervised learning – Back-propagation

Probabilistic nets

In the following slides we will give a very brief introduction to neural networks. We will discuss only two network types – a feedforward net trained with the back-propagation learning algorithm, and a probabilistic net. We made these two choices since the back-prop net is perhaps the most widely used network to describe the working of neural nets and the probabilistic net is closely related to the probabilistic ideas we discussed in the pattern recognition presentation.



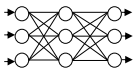
What is a Neural Network?

According to the *DARPA Neural Network Study*
(1988, AFCEA International Press, p. 60):

... a neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes.

Here is a formal definition of a neural network. Generally speaking, a neural net has many interconnected elements which work in unison to accomplish a particular task.

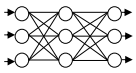
Characteristics of Neural Nets



The good news: They exhibit some brain-like behaviors that are difficult to program directly like:
learning
association
categorization
generalization
feature extraction
optimization
noise immunity

The bad news: neural nets are:
black boxes
difficult to train in some cases

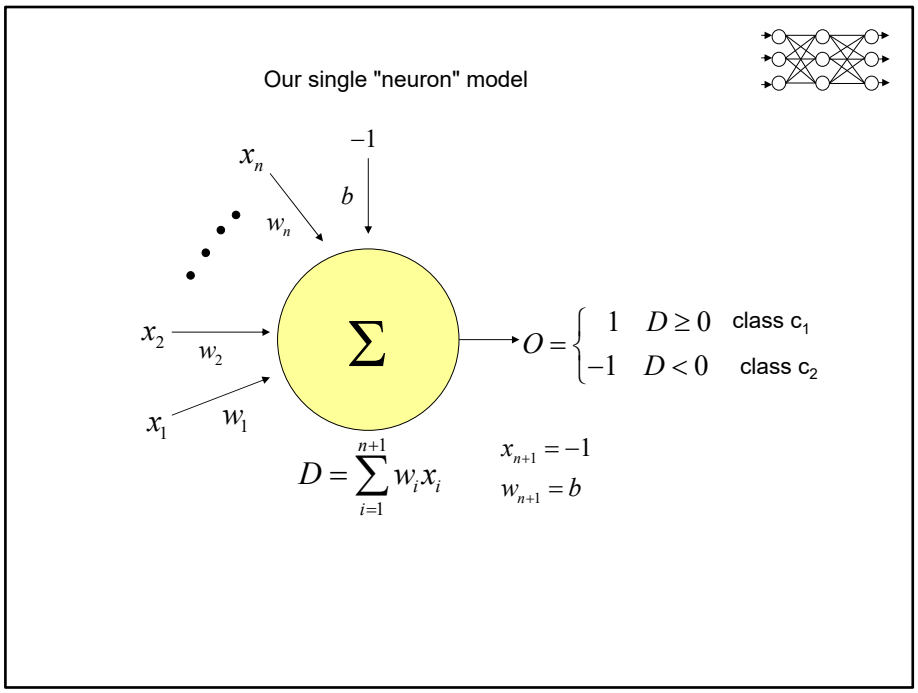
Neural nets have some very good characteristics but also some bad ones as well. The power exhibited by neural nets generally far outweigh their negative aspects.



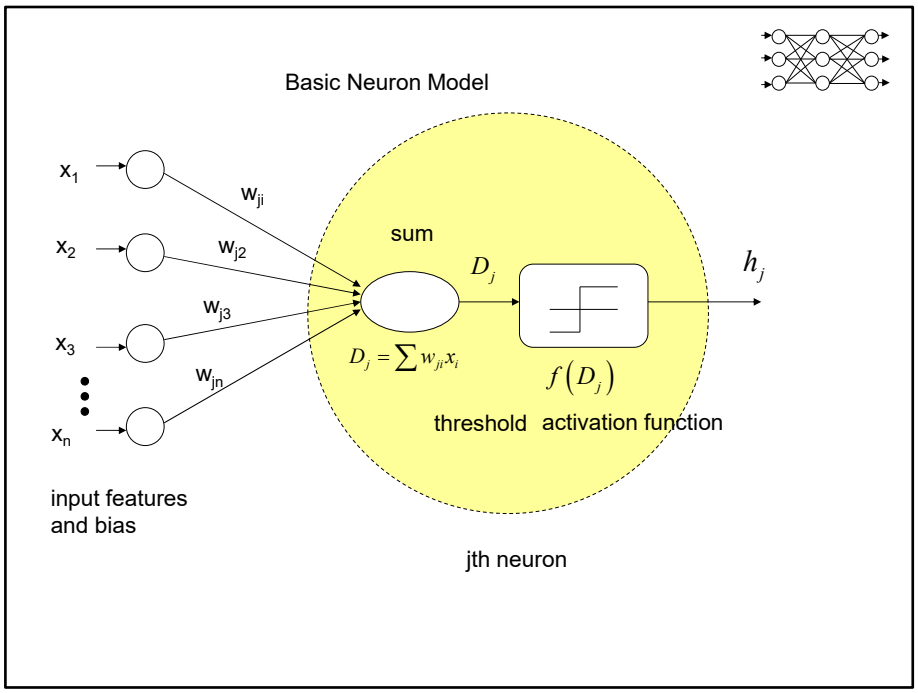
There is a wide range of neural network architectures:

Multi-Layer Perceptron (Back-Prop Nets) 1974-85
Neocognitron 1978-84
Adaptive Resonance Theory (ART) 1976-86
Sef-Organizing Map 1982
Hopfield 1982
Bi-directional Associative Memory 1985
Boltzmann/Cauchy Machine 1985
Counterpropagation 1986
Radial Basis Function 1988
Probabilistic Neural Network 1988
General Regression Neural Network 1991
Support Vector Machine 1995

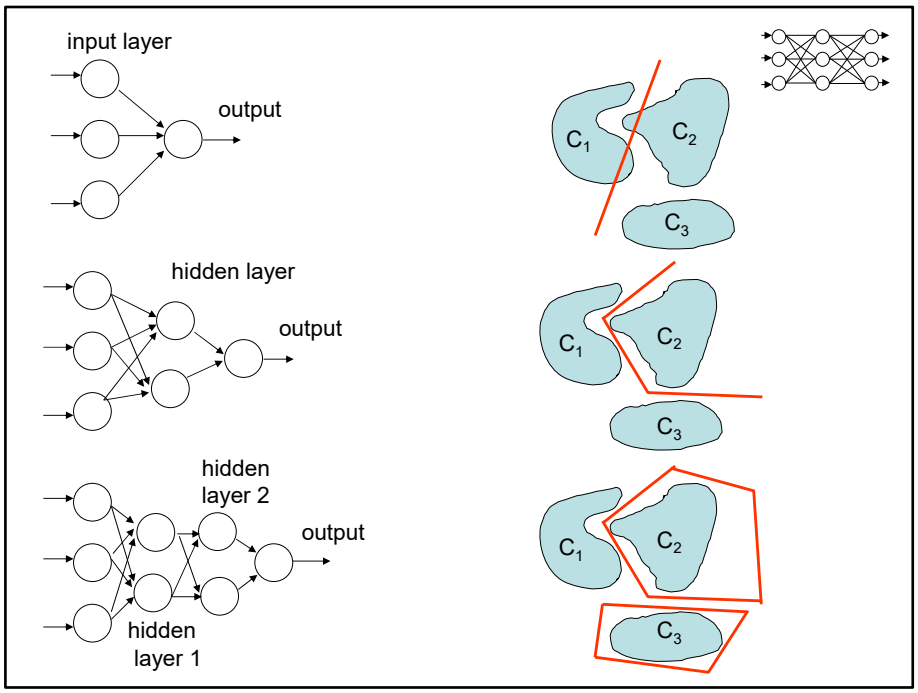
Here we show a number of different neural nets developed in the 80s and 90s. We will only describe back-propagation nets and probabilistic nets, as mentioned earlier.



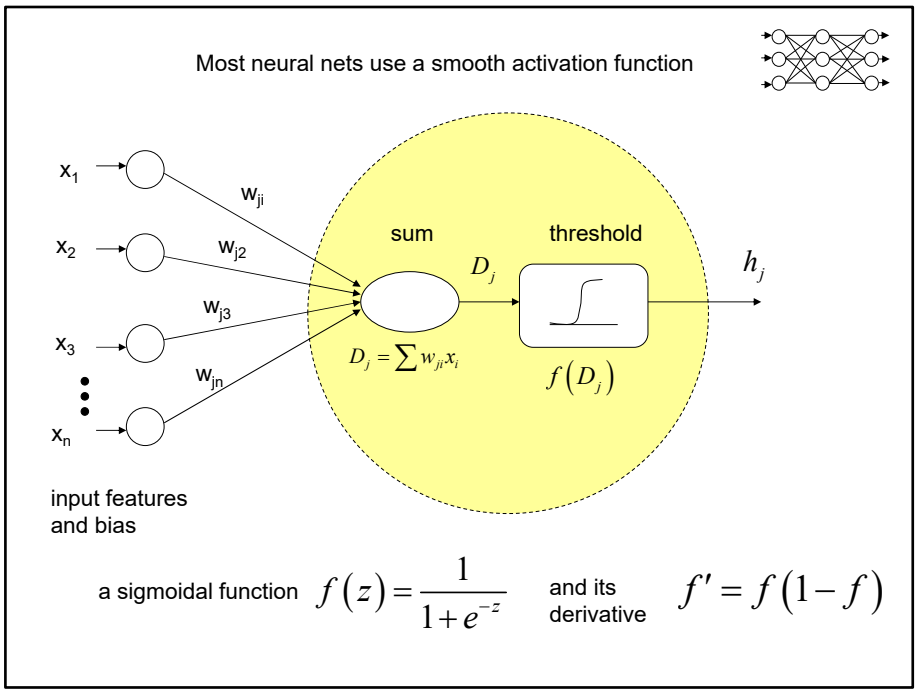
Here is the model of a single “neuron” that implements the linearly separable classifier we discussed in the pattern recognition notes.



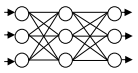
Here is a more general “neuron” model that takes a set of weighted inputs, sums those inputs, and then applies an “activation function” to that sum to determine an output.



We saw that with a single neuron we can generate a linearly separable classifier. By adding one or more layers we can generate a network of neurons that can handle more complex problems since the decision surfaces (discriminants) are not limited to a single hyperplane. These decision surfaces are defined by the weights in the connections between neurons so a key aspect is how we learn to choose those weights.



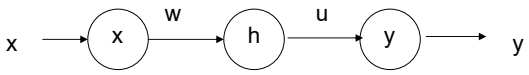
Neural nets usually use a smooth activation function that we will see facilitates the determination of the network weights. A sigmoidal function is a popular choice.



Major question – How do we adjust the weights to learn the mapping between inputs and outputs?

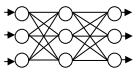
Answer: Use the back propagation algorithm, which is just an application of the chain rule of differential calculus.

Consider this simple example



$$y = f(u\,h)$$
$$h = f(w\,x)$$
$$y = f(u\,f(w\,x))$$
$$= F(u, w, x)$$

The crucial question is how to determine the weights. A procedure, called the **back propagation algorithm**, allows us to answer that question. To see how back propagation works, consider a simple case of a set of three neurons connected together with unknown weights, w, u.



To learn the weights we can try to minimize the output error

i.e. , we start with an initial guess for the weights and then present an example of known input x and output value d (training example). Then we form up the error

$$E = \frac{1}{2}(d - y)^2$$

and adjust the weights to reduce this error. Since

$$\Delta E = \frac{\partial E}{\partial u}\Delta u + \frac{\partial E}{\partial w}\Delta w$$

we can make $\Delta E \leq 0$

by choosing

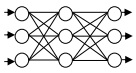
$$\frac{\partial E}{\partial u} = -\lambda \Delta u$$

$$\frac{\partial E}{\partial w} = -\lambda \Delta w$$

λ … constant

To determine the weights we take an example where an input x determines a known output d and try to minimize the output error of the network on this example. We can make the error always smaller by making the derivatives of the error function with respect to the weights proportional to the change of weights themselves as shown.

This leads to the adjustment rule



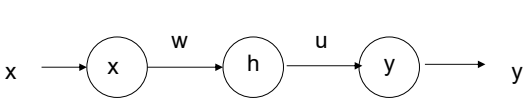
$$u(m+1) = u(m) - \mu \frac{\partial E}{\partial u(m)}$$

$$w(m+1) = w(m) - \mu \frac{\partial E}{\partial w(m)}$$

μ … learning rate

So we now need to find these derivatives of E with respect to the weights.

This can be written as a rule for changing the weights in terms of the derivatives of the error.



$$y = f(u\,h)$$
$$h = f(w\,x)$$

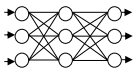
$$y = f(u\,f(w\,x))$$
$$= F(u, w, x)$$

For the derivatives involving the weight between the hidden layer and output layer

$$\frac{\partial E}{\partial u(m)} = \frac{\partial}{\partial u}\left[\frac{1}{2}(d-y)^2\right] = \frac{\partial}{\partial y}\left[\frac{1}{2}(d-y)^2\right]\frac{\partial y}{\partial u}$$
$$= (y-d)\,f'(u\,h)\,h = (y-d)\,y(1-y)\,h$$
$$= (y-d)\,y(1-y)\,h\Big|_{u(m),w(m)}$$

Since the relationship between the network output and input is just a composite function, we can obtain the error derivatives explicitly as shown here for one error derivative, ∂E/∂u (the expression for the derivative shown earlier for the sigmoidal function is used here). Choosing a smooth function like the sigmoid allows us to compute these derivatives.

Similarly for the weight between the hidden layer and input layer



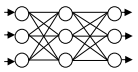
$$\frac{\partial E}{\partial w(m)} = \frac{\partial}{\partial w}\left[\frac{1}{2}(d-y)^2\right] = \frac{\partial}{\partial y}\left[\frac{1}{2}(d-y)^2\right]\frac{\partial y}{\partial w}$$

$$= (y-d)\, f'(u\,h)\frac{\partial(u\,h)}{\partial w} = (y-d)\, y(1-y)\, u\, h'(w\,x)\, x$$

$$= (y-d)\, y(1-y)\, h(1-h)\, ux\Big|_{u(m),w(m)}$$

Here is the other error derivative. Thus, we are really just applying the chain rule of Calculus in this backpropagation algorithm.

Thus, the training algorithm is:

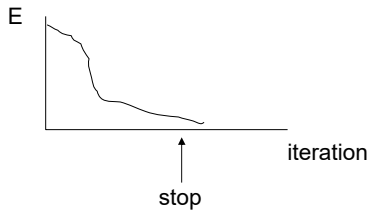


1. Initialize weights to small random values
2. Using a training set of known pairs of inputs and outputs (x, d) change the weights according to

$$w(m+1) = w(m) - \mu(y-d)\,y(1-y)\,h(1-h)\,u\,x\big|_{u(m),w(m)}$$

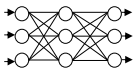
$$u(m+1) = u(m) - \mu(y-d)\,y(1-y)\,h\big|_{u(m),w(m)}$$

with $y = f(u\,h)$, $h = f(w\,x)$ until $E = \frac{1}{2}(d-y)^2$ becomes sufficiently small

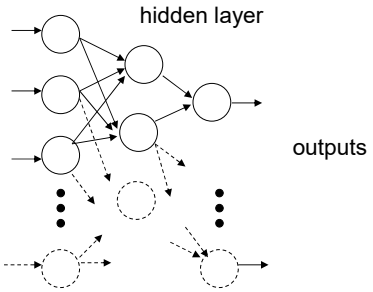


This is an example of supervised learning

Thus, we now know how to use a set of known examples of input/output combinations (called a training set) to change the weights to minimize the output error. Once the error is small enough, we can say the neural network has been trained. This is called **supervised learning**.



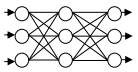
One of the most popular neural nets is a feed forward net with one hidden layer trained by the back propagation algorithm

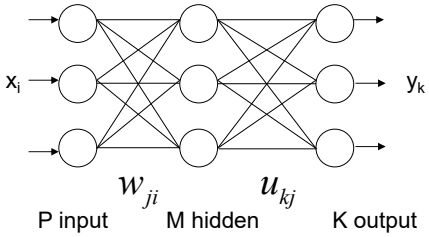


It can be shown that in principle this type of network can represent an arbitrary input-output mapping or solve an arbitrary classification problem

Now, consider a feed forward network with a hidden layer.

Back propagation algorithm (three layer feed forward network)





$$u_{kj}^{new} = u_{kj}^{old} - \mu\left(y_k - d_k\right) y_k \left(1 - y_k\right) h_j \quad \left(k = 1,...,K \;\; j = 1,...,M\right)$$

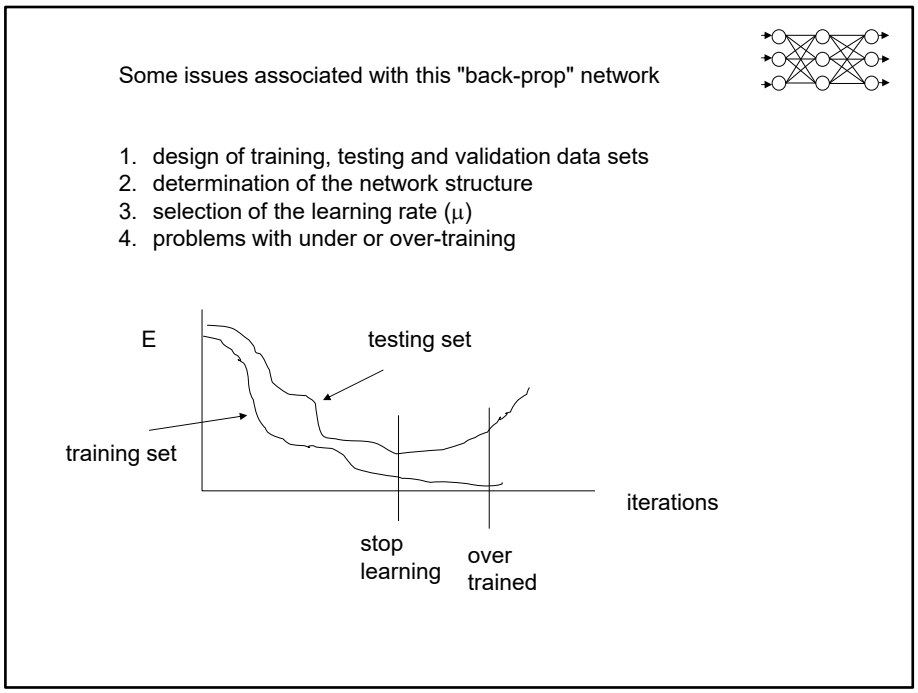
$$w_{ji}^{new} = w_{ji}^{old} - \mu \sum_{k=1}^{K} \left(y_k - d_k\right) y_k \left(1 - y_k\right) u_{kj}^{old} \; h_j \left(1 - h_j\right) x_i \quad \left(j = 1,...,M \;\; i = 1,...,P\right)$$

with

$$y_k = \frac{1}{1 + \exp\left(-\sum_{j=1}^{M} u_{kj}^{old} h_j\right)}$$

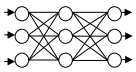
$$h_j = \frac{1}{1 + \exp\left(-\sum_{i=1}^{P} w_{ji}^{old} x_i\right)}$$

We can apply the same backpropagation algorithm to train this neural network.



Here are some of the questions that occur with such training of such "**back-prop neural nets".** For example, how do we choose sets of training examples and examples to validate the performance? Also, how do we choose the actual network architecture (number of nodes, layers)? There is a **learning rate**, μ , in the backpropagation algorithm, that also must be chosen. Also, we need to determine when to stop the training since we may make the error of the network on the training examples be very small, but a smaller training error does not always mean a better performance of the network on other examples, as we may simply have trained the network to only recognize the testing examples very well.

Some important issues for neural networks:

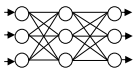


Pre-processing the data to provide:

- reduction of data dimensionality
- noise filtering or suppression
- enhancement
  - strengthening of relevant features
  - centering data within a sensory aperture or window
  - scanning a window over the data
- invariance in the measurement space to:
  - translations
  - rotations
  - scale changes
  - distortion
- data preparation
  - analog to digital conversion
  - data scaling
  - data normalization
  - thresholding

Here are some issues of preprocessing the input data (training examples) that might be useful.

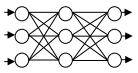
Some examples of pre-processing include



1-D and 2-D FFTs
Filtering
Convolution Kernels
Correlation Masks or Template Matching
Autocorrelation
Edge Detection and Enhancement
Morphological Image Processing
Fourier Descriptors
Walsh, Hadamard, Cosine, Hartley, Hotelling, Hough Transforms
Higher order spectra
Homomorphic Transformations (e.g. Cepstrums)
Time-Frequency transforms (Wavelet, Wigner-Ville, Zak)
Linear Predictive Coding
Principal Component Analysis
Independent Component Analysis
Geometric Moments
Thresholding
Data Sampling
Scanning

Here are some explicit pre-processing steps. A feedforward neural network trained with the back propagation algorithm used in conjunction with such steps can be a very powerful tool.

Probabilistic Neural Network (PNN)



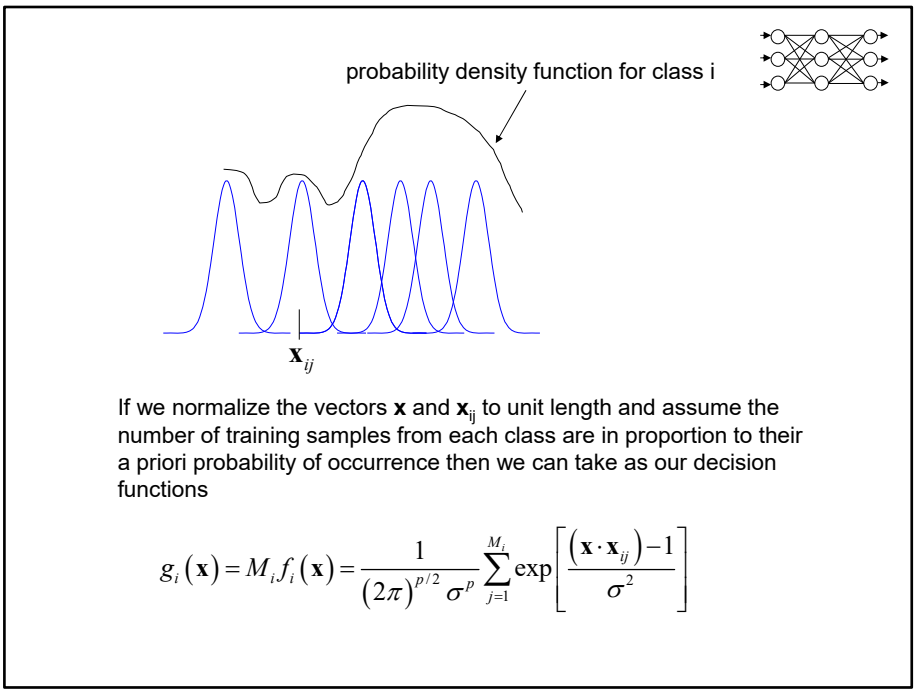
Basic idea:
Use training samples themselves to obtain a representation of the probability distributions for each class and then use Bayes decision rule to make a classification

Basis functions usually chosen are Gaussians:

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\sigma^p M_i} \sum_{j=1}^{M_i} \exp\left[\frac{-(\mathbf{x}-\mathbf{x}_{ij})^T(\mathbf{x}-\mathbf{x}_{ij})}{2\sigma^2}\right]$$

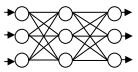
i … class number (i = 1, 2, …, N)
j … training pattern number
$\mathbf{x}_{ij}$ … j th training pattern from i th class
M_i … number of training vectors in class i
p .. dimension of vector **x**
$f_i(\mathbf{x})$ = sum of Gaussians centered at each training pattern from the ith class to represent the probability density of that class
σ … smoothing factor (standard deviation, width of the Gaussians)

Here is another neural network, called a probabilistic neural net, that uses the samples (training examples) directly to define the probability distributions for a set of output classes and then applies Bayes decision rule to make a classification. Gaussian functions are often used, as shown here.



Here are the decision functions that can be defined from the training a samples.

$$g_i(\mathbf{x}) = M_i f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\sigma^p} \sum_{j=1}^{M_i} \exp\left[\frac{(\mathbf{x}\cdot\mathbf{x}_{ij})-1}{\sigma^2}\right]$$



Since we decide for a given class k based on

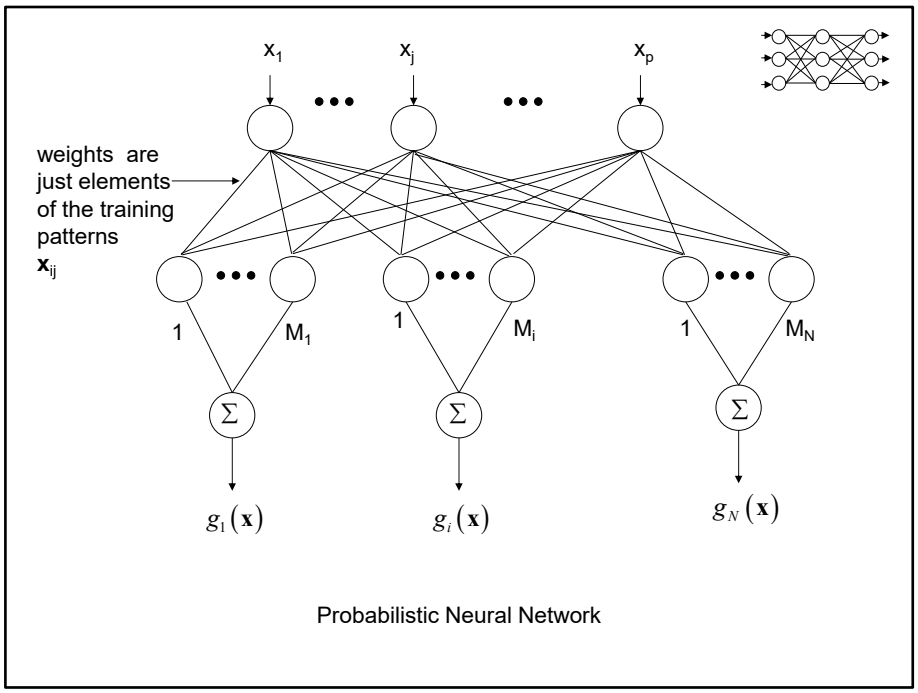
$g_k(\mathbf{x}) > g_i(\mathbf{x})$ for all $i = 1, 2, ..., N \quad (i \neq k)$

the common constant outside the sum makes no difference and we can take

$$g_i(\mathbf{x}) = \sum_{j=1}^{M_i} \exp\left[\frac{(\mathbf{x}\cdot\mathbf{x}_{ij})-1}{\sigma^2}\right]$$

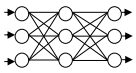
This can now be easily implemented in a neural network form

Then we simply choose the largest decision function to make a classification decision.



This whole process can be structures as a neural network where the weights are now simply the known elements of the training examples.

Characteristics of the PNN



1. no training, weights are just the training vectors themselves

2. only parameter that needs to be found is the smoothing factor, σ

3. outputs are representative of probabilities of each class directly

4. the decision surfaces are guaranteed to approach the Bayes optimal boundaries as the number of training samples grows

5. "outliers" are tolerated

6. sparse samples are adequate for good network performance

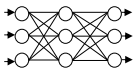
7. can update the network as new training samples become available

8. needs to store all the training samples, requiring a large memory

9. testing can be slower than with other nets

Here are some of the properties of the probabilistic neural network. We see that it is a nice illustration of the probabilistic pattern recognition ideas we presented earlier in a neural network form.

References



Specht, D.F., " Probabilistic neural networks," Neural Networks, 3, 109-118, 1990.

Haykin, S., *Neural Networks, a Comprehensive Foundation*, 2nd Ed., Prentice-Hall, 1999.

Haykin, S., Neural Networks and Learning Machines, 3rd Ed., Pearson, 2016.

Bishop, C.M., *Neural Networks for Pattern Recognition*, Clarendon Press, 1995.

Bishop, C.M., Pattern Recognition and Machine Learning, Springer, 2006.

Geron, A., Hands-on Machine Learning with SciKit-Learn, Keras, and TensorFlow, 2nd Ed., O Reilly Media, 2019.

Here are a few references that contain more information on back-prop nets and probabilistic nets, as well as others.