University of Alberta

Top-Orthogonal-to-Bottom-Electrode Capacitive Micromachined Ultrasound Transducers for Low-Cost 3D Ultrasound Imaging

by

Alexander Sampaleanu

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Microsystems and Nanodevices

Electrical and Computer Engineering

©Alexander Sampaleanu
Fall 2013
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Abstract

Capacitive micromachined ultrasound transducers (CMUTs) offer a promising solution for three-dimensional ultrasound imaging due to their ability to be easily fabricated in two-dimensional arrays. However, channel routing is still non-trivial and cost-intensive. The subject of this thesis is a 'Top-Orthogonal-to-Bottom-Electrode' (TOBE) 2-D CMUT array architecture, along with its respective row-column addressing schemes, which can significantly reduce the channel count of a 2-D array, such that only 2N channels are required for an N x N array. Background on acoustics, ultrasound imaging, and CMUT technology is presented before the novel array architecture and imaging schemes is described. Arrays were fabricated using a sacrificial release method, and acoustic immersion and air-coupled tests, along with imaging simulations, were done to validate array functionality. Results showed that single-element actuation is possible with a row-column addressing scheme, allowing for novel imaging schemes which could result in significantly more cost-effective 3-D ultrasound imaging.

Acknowledgements

I would like to thank the many people without whom completing this thesis and this degree would not have been possible. First and foremost I would like to thank my supervisor Dr. Roger Zemp, who came to me at the beginning and offered me the opportunity to start my degree working on an exciting and challenging project. The constant support, guidance, and knowledge provided throughout the degree were invaluable and irreplaceable. I believe there are few supervisors who would demonstrate such consistent dedication and motivation to their students, and whose passion and excitement for the subject matter was as pronounced as his. These qualities made it that much easier for me to get motivated about the work as well, even through the setbacks and barriers that inevitably accompany research.

I would also like to acknowledge the many other members of the University Of Alberta Department Of Electrical Engineering for their help and guidance through my degree. I would like to thank the NanoFab staff for offering their training and guidance to me when using the fabrication equipment, from which valuable experience was gained, especially the training coordinators: Scott Munro, Les Schowalter, Jolene Jalota, and Stephanie Bozic. I acknowledge all the professors that helped get me through my classes and the administrative and technical support staff for all the little things such as ordering equipment, machining necessary components, and helping me navigate through all the paperwork that always seems to require one more piece of information. I would also like to give mention to the other members of my lab group for offering engaging discussions on all manner of topics, giving moral and technical support when needed, and always reminding me to think twice and to live life to its fullest outside of research. Specifically I would mention Peiyu Zhang, Abhijeet Kshirsagar, Parsin Hajireza, Tyler Harrison, and Ryan Chee for their specific help with the research. In addition, I would like to acknowledge the students from Dr. Walied Moussa's lab helping us use their test equipment.

I would also like to acknowledge the Faculty of Graduate Studies and Research, the National Sciences and Engineering Research Council of Canada, and Alberta Innovates for providing much of the funding which made this degree possible. Furthermore, I would like to thank Micralyne Inc. and Microfab Service GmbH for their assistance with CMUT fabrication.

Last, but certainly not least, I would like to thank my parents for always offering their support in my goals. Despite being separated by hundreds of kilometers, I always felt like they were just a short ways away, eager to help me whenever I needed it.

Table of Contents

Chapter 1: Introduction **1**
1.1 Goals and Incentives 1
1.2 Objectives of Thesis 2
1.3 Organization of Thesis 3

Chapter 2: Background to Ultrasound **4**
2.1 Introduction to Ultrasound Technology 4
2.1.1 Ultrasound Wave Basics 4
2.1.2 Acoustic Wave Propagation 5
2.1.3 Wave Properties at Interfaces 6
2.1.4 Ultrasound Scattering and Absorption 10
2.2 Ultrasound Field Profile Analysis 11
2.2.1 Analytical Methods for Calculating Ultrasound Fields 11
2.2.2 Approximations for Calculating Ultrasound Fields 12
2.3. Piezoelectric Ultrasound Transducer Technology 13

Chapter 3: Ultrasound Imaging 15
3.1 Pulse-echo Pressure Field Analysis 15
3.2 Linear Array Imaging 17
3.2.1 Introductory Imaging Concepts 17
3.2.2 Linear Array Transducer Pressure Field Analysis 18
3.2.3 Beamforming with Linear Arrays 21
3.3 Three-dimensional Ultrasound Imaging 24
3.3.1 Three-dimensional Ultrasound Imaging using Linear Arrays 25
3.3.2 Two-dimensional Planar Arrays 25
3.3.3 Alternative 3-D Imaging Solutions 28
3.3.4 Sparse Arrays 29
3.4 Synthetic Aperture Imaging 32

3.4.1 Synthetic Aperture Imaging Basics 32
3.4.2 Synthetic Aperture Imaging for 2-D Arrays 35

Chapter 4: Capacitive Micromachines Ultrasound Transducers 36
4.1 Introduction to CMUTs 36
4.1.1 CMUT Structure 36
4.1.2 CMUT Operation 37
4.1.3 CMUT Parameters and Equivalent Electrical Model 39
4.1.4 CMUT Operating Modes 43
4.2 Comparisons with Piezoelectric Transducers 44
4.2.1 Advantages of CMUT Technology 44
4.2.2 Disadvantages of CMUT Technology 46
4.3 CMUT Fabrication 46
4.3.1 Sacrificial Release Process 47
4.3.2 Sacrificial Release Process Variations 50
4.3.3 Sacrificial Release Fabrication Observations and Results 51
4.3.4 Wafer-bonding Process 53
4.3.5 Wafer-bonding Process Variations 55
4.3.6 Process Comparisons 55
4.4 CMOS Integration and Packaging 55

Chapter 5: Top-Orthogonal-to-Bottom-Electrode (TOBE) Arrays 59
5.1 Introduction to TOBE Arrays 59
5.1.1 TOBE Array Structure 59
5.1.2 TOBE Array Fabrication Process 60
5.2 Imaging Using TOBE Arrays 61
5.2.1 Imaging Scheme 1 61
5.2.2 Imaging Scheme 2 62
5.2.3 Imaging Scheme 3 64
5.2.4 Imaging Simulations 65
5.3 Feasibility Test Results 68

5.3.1 RF Actuation Vibrometer Tests in Air Operation 68
5.3.2 Acoustic Immersion Transmit Testing 72
5.3.3 Acoustic Immersion Receive Testing 74
5.4 Interface Electronics...76
5.5 Discussion of Results 77

Chapter 6: Conclusion..80
6.1 Summary 80
6.2 Future Directions...81

References...83

Appendix 1: Synthetic Aperture 3D Ultrasound Imaging Schemes witih S-Sequence Bias-Encoded Top-Orthogonal-to-Bottom-Electrode 2D CMUT Arrays..88

List of Tables

5.1 SNR comparisons between various 3-D imaging schemes..............................67

5.2 Collapse-mode TOBE CMUT operation and cross-talk comparison 72

List of Figures

Figure 2.1: Illustration of four types of waves important to ultrasound propagation. a) Longitudinal ; b) Transverse ; c) Lamb ; d) Stoneley 5

Figure 2.2: Behaviour of a wave at the interface between two media 7

Figure 2.3: Depiction of a matching layer between the ultrasound transducer and the medium...9

Figure 2.4: Effect of frequency dependent attenuation on signal shape 11

Figure 3.1: A typical transmit-receive ultrasound setup.......................................16

Figure 3.2: a) A-scan b) B-mode image constructed from multiple B-scans 17

Figure 3.3: A linear array transducer 18

Figure 3.4: Pressure field profile for a linear array transducer 19

Figure 3.5: Pressure field profiles demonstrating the effect of changing linear array parameters. a) The array has the same properties as the one given in figure 3.4 as a base comparison. b)The frequency has been doubled. c)The frequency has been reduced by 2.5 times. d) The array is the same size, but has fewer elements and an increased kerf. e) The kerf has been increased but the number of elements is the same as in a), resulting in a larger aperture. f)The number of elements has decreased, thus decreasing the aperture size, but all other parameters remained constant. 21

Figure 3.6: Pressure field profile from a linear transducer with a) no steering, and b) 75° steering 24

Figure 3.7: A 2-D plane array transducer 26

Figure 3.8: Beamforming arrangements for a) Linear array transducer; b)1.25-D transducer; c) 1.5-D Transducer...29

Figure 3.9: Examples of sparse 2-D array designs: a)Vernier (periodic) sparse transmit array; b)Vernier (periodic) sparse receive array; c)Random sparse transmit array; d)Random sparse receive array..31

Figure 3.10: Visualization of the monostatic synthetic aperture imaging approach 33

Figure 3.11: Visualization of the synthetic transmit aperture imaging method.....34

Figure 4.1: Diagram of a typical CMUT cell..37

Figure 4.2: Series of capacitance-voltage curves of one of our CMUTs...............39

Figure 4.3: a) Mason equivalent model of a CMUT when transmitting ultrasound. b) Mason equivalent model of a CMUT when receiving ultraousnd. $R_{s:}$ Source impedance; L_{TUN}: Tuning inductance; C_P, $R_{P,}$: Parasitic capacitances and resistances; C_0: CMUT capacitance; n: Electromechanical conversion factor; Z_{mem}: CMUT Membrane Impedance; Z_{med}: Impedance of transmission medium; R_{loss}: Mechanical loss; C_{amp}: Amplifier input capacitance; R_{amp}: Amplifier input resistance..................40

Figure 4.4: Electromechanical efficiency vs. voltage curve for one of our CMUTs..42

Figure 4.5: Operation of a CMUT in pre-collapse mode on the left, and collapse mode on the right 43

Figure 4.6: SEM image of a dual-frequency CMUT array...................................45

Figure 4.7: Process flow for a sacrificial release CMUT fabrication process: a) Deposition of insulating layer; b) Deposition and patterning of initial sacrificial layer; c) Deposition and patterning of second sacrificial layer to define cavity; d) Deposition of membrane layer; e) Etching of etch holes f) Sacrificial release step to release membrane; g) Deposition and patterning of sealing plugs; h) Deposition and patterning of metallization layer after etching access holes 48

Figure 4.8: Diagram depicting the consequence of poor etch selectivity during sacrificial layer etch 49

Figure 4.9: Process flow for a simple wafer-bonding fabrication method: a) Thermal oxide is grown on a silicon wafer; b) CMUT cavity is etched in the oxide; c) Thermal oxide is re-grown to act as insulation layer; d) SOI wafer is fusion bonded to Si wafer; e) Handle and buried oxide are removed; f) Metallization layer is deposited and patterned; g) Additional passivating layer is deposited if necessary...54

Figure 4.10: SOI wafer with trenches etched into silicon device layer. The silicon device layer is heavily doped to act as a bottom electrode. 55

Figure 4.11: Diagram of a post-CMOS integrated CMUT structure.....................57

Figure 4.12: Diagram of a flip-chip bonded CMOS-integrated CMUT structure (CMUT structures are shown with a wafer-bonded architecture)..............58

Figure 5.1: 3-D model of our fabricated TOBE array 60

Figure 5.2: SEM image of a single 2 x 2 element in one of our TOBE arrays. 61

Figure 5.3: Imaging Scheme 1: Orthogonal one-way focused synthetic aperture imaging 62

Figure 5.4: Imaging Scheme 2: Orthogonal two-way actively focused synthetic aperture imaging..........63

Figure 5.5: Visual representation of the S-sequence aperture encoded imaging method..........65

Figure 5.6: Cross-Range-Maximum-Amplitude point-spread function plots comparing three different imaging schemes with a fully-wired 2-D phased array in: a) the lateral direction and b) the elevational direction. 66

Figure 5.7: Noise-added C-scan point spread functions of: a) S-sequence aperture encoding scheme; b) 2-D phased array; c) Scheme 1; d) Scheme 2; e) Identity matrix aperture encoding scheme. 67

Figure 5.8: Scheme 1 test set-up: all columns and rows are grounded except for column 1 and row 1 (which is floating) 69

Figure 5.9: Dynamic displacement vibrometer measurements of a 4 x 4 section of a TOBE CMUT array operating in scheme 1..........69

Figure 5.10 Linear array test set-up for comparing crosstalk. 70

Figure 5.11: Single-element actuation test set-up: all columns and rows aregrounded except for column 1, which is biased, and row 1, which has signal applied. 70

Figure 5.12: Dynamic displacement vibrometer measurements of a 4 x 4 section of a TOBE CMUT array operating in scheme 2 with: a) 0V bias on column 1; b) 70V bias on column 1 71

Figure 5.13: a) Maximum recorded pressure signal from TOBE array; b) Fourier transform of pressure signal..........73

Figure 5.14: a)Near-field x- and y-direction pressure variations of TOBE array. b) Pressure field map of entire TOBE array superimposed on SEM image of TOBE CMUT die. 74

Figure 5.15: a) Maximum recorded received signal from TOBE array at 100V bias; b) Change in maximum received signal amplitude due to variations in bias applied bias voltage 75

Figure 5.16: Received signal variations of TOBE array due to translation of a focused transducer away from active area. 76

List of Symbols and Abbreviations

List of Symbols

Symbols

A:	CMUT membrane area
c_0:	Propagation speed of an ultrasound wave
c_L:	Propagation speed of a longitudinal wave in a solid medium
c_T:	Propagation speed of a transverse wave in a solid medium
C:	Capacitance
d:	Pitch of a linear transducer array (Chapter 3) ; Initial CMUT gap height (Chapter 4)
d_x:	Pitch of a 2-D transducer array in the lateral direction
d_y:	Pitch of a 2-D transducer array in the elevational direction
D:	Total width of a linear transducer array (Chapter 3) ; Flexural rigidity (Chapter 4)
D_x:	Size of a 2-D array in the lateral direction
D_y:	Size of a 2-D array in the elevational direction
E:	Young's modulus of a solid medium
f:	Frequency
F_e:	Electrostatic force
F_s:	Mechanical restoring force
$h_r(\boldsymbol{r}, t)$:	Impulse response function of some source point at a point **r** on a transducer
$h_t(\boldsymbol{r}: t)$:	Impulse response function of a transducer at a point **r** in space
H:	Height of a transducer element
k:	Wave number of an ultrasound wave
k_f:	Kerf
k_s:	Spring constant
k_x:	Spatial frequency component in the x-direction

k_y: Spatial frequency component in the y-direction

k_z: Spatial frequency component in the z-direction

I: Intensity

N: Number of elements in a transducer array

p: ressure

r: A point in space **r** = (x,y,z)

R: The distance between an observation point and some source point

R_I: Intensity reflection coefficient

R_L: Longitudinal wave intensity reflection coefficient at a fluid-solid interface

R_N: Normal incidence intensity reflection coefficient

t: Time

t_m: CMUT membrane thickness

T_I: Intensity transmission coefficient

T_L: Longitudinal wave intensity transmission coefficient at a fluid-solid interface

T_N: Normal incidence intensity transmission coefficient

T_T: Transverse wave intensity transmission coefficient at a fluid-solid interface

$\boldsymbol{v_L}$: Velocity of an ultrasound wave propagating in a fluid

$v_{no}(t)$: Time-varying normal velocity of a transducer surface

V: Voltage

V_c: Collapse Voltage

V_{SB}: Snapback Voltage

W: Width of a transducer element

x_i: X-coordinate on a transducer surface plane

x_o: X-coordinate of an observation point

y_i: Y-coordinate on a transducer surface plane

y_o: Y-coordinate of an observation point

z_i: Z-coordinate on a transducer surface plane

z_o: Z-coordinate of an observation point

Z_0: Characteristic acoustic impedance of a fluid medium

Z_L: Characteristic acoustic impedance of a solid medium for longitudinal waves

Z_T: Characteristic acoustic impedance of a solid medium for transverse waves

Greek Symbols

α: Total amplitude attenuation coefficient

α_0: Attenuation constant

α_a: Amplitude attenuation coefficient due to absorption

α_S: Amplitude attenuation coefficient due to scattering

ε: Dielectric permittivity

θ_c: Critical angle at which the transmitted angle at an interface is 90 degrees

θ_i: Incidence angle of a wave on an interface between two media

θ_{int}: Intromission angle at which the all energy of a wave incident on an interface between two media is transmitted

θ_L: Angle of the longitudinal wave transmitted through the interface between a fluid-solid interface

θ_r: Angle of the wave reflected at an interface between two media

θ_s: Polar steering angle

θ_t: Angle of the wave transmitted through the interface between two fluid media

θ_T: Angle of the transverse wave transmitted through the interface between a fluid-solid interface

κ: Adiabatic compressibility of a fluid medium

λ: Wavelength

μ: Shear viscosity of a fluid medium

μ_B: Bulk viscosity of a fluid medium

$\xi_o(x,y)$: Spatial variation of a transducer surface velocity

ρ_0: Equilibrium density of a medium

σ: Poisson's ratio of a solid medium

τ_n: Element delay for *nth* element from center of array.

ϕ: Velocity potential of an ultrasound wave

ϕ_s: Beam steering angle

φ_s: Azimuth beam steering angle

ω: Angular frequency

$\Omega(x_i, y_i)$: Aperture function of a transducer

Other Symbols

∇: Gradient of

List of Abbreviations

BOE: Buffered Oxide Etch

CMUT: Capacitive Micromachined Ultrasound Transducer

CVD: Chemical Vapour Deposition

CT: Computed Tomography

EDP: Aqueous Ethylene Diamine and Pyrocatechol (used for etching silicon)

FFT: Fast Fourier Transform

FWHM: Full-width-half-maximum

HNA: Combination of HF/Nitric Acid (used for etching silicon)

ICPCVD: Inductively Coupled Plasma Chemical Vapour Deposition

ISLR: Integrated Sidelobe Ratio

LPCVD: Low-pressure Chemical Vapour Deposition

MEMS: Microelectromechanical Systems

MRI: Magnetic Resonance Imaging

NDT: Non-destructive Testing

NEP: Noise Equivalent Pressure

PECVD: Plasma-enhanced Chemical Vapour Deposition

R_x: Receive

RIE: Reactive Ion Etching

SAI: Synthetic Aperture Imaging

SEM: Scanning Electron Microscope

SNR: Signal-to-noise Ratio

SRAI: Synthetic Receive Aperture Imaging

STAI: Synthetic Transmit Aperture Imaging

TEOS: Tetraethyl Orthosilicate

TOBE: Top-orthogonal-to-bottom-electrode (refers to architecture of CMUT array)

T_x: Transmit

Chapter 1

Introduction

1.1 Goals and Incentives

The use of ultrasound and ultrasonic transducers has become increasingly prevalent in many different fields and with numerous unique applications. In industrial applications, it is used frequently in non-destructive testing (NDT) to characterize different materials and test them for flaws. Industries that use this technique include the oil and gas, transportation, and aerospace industries among others. Other uses in industry include object detection (for security or proximity detection), industrial cleaning, and flow sensing, which can determine the flow rate of a specific fluid, such as natural gas in a pipeline. Despite all these aforementioned applications, perhaps the most prevalent and well-known application of ultrasound is in the medical field, especially as a medical imaging tool. In this field, ultrasound imaging is attractive because, compared to other medical imaging modalities such as computed tomography (CT) or magnetic resonance imaging (MRI), it is relatively cheap and does not require the complicated setup that other imaging technologies do. Some ultrasound imaging systems are even portable, allowing for their use outside of a clinic or hospital environment. Ultrasonic imaging can also be more conducive to patient comfort since it is non-invasive and does not use any potentially harmful, ionizing radiation. It is also endearing in its ability to provide real-time images, and, in addition to imaging, ultrasound can also be used for therapeutic applications to treat problems such as prostate cancer [1]. Overall, ultrasound is an extremely important and irreplaceable tool in the medical field, having a market size of over five billion annually in 2009 [2], with substantial potential for further development.

Given the demand for ultrasound technology in so many different fields, the advantages of further improving and developing it are undeniable. Although the technology has already improved significantly since it was introduced in the early 90s, much of the research and study on ultrasound transducers was focused on piezoelectric technology (this technology is described more in section 2.3). Over the past two decades, however, an increasing amount of research has been devoted to a new type of ultrasound transducer technology known as capacitive micromachined ultrasound transducers (CMUTs), which use electrostatic actuation instead of piezoelectric actuation. Further analysis on the principles behind this technology are provided in chapter 4, but in brief, in addition to having improved performance over piezoelectric transducers, CMUTs are fabricated using silicon-based MEMS micromachining techniques, allowing for large-scale, cost-effective fabrication. These techniques also offer increased flexibility in terms of ultrasound array design and electronics integration. The ability to replace the delicate, manual labour required to fabricate piezoelectric transducers with simple, high-yield batch-process techniques such as optical

lithography allow CMUT arrays to be much smaller or have complicated arrangements of elements. One specific type of array where the CMUT fabrication process offers the greatest advantages is the 2-D ultrasound array, which can be used to perform 3-D ultrasound imaging. This imaging method is desirable due to its ability to provide higher quality and more informative volumetric images that require less subjective analysis compared to 2-D images. To maintain a good imaging quality, a large number of transducers must be very tightly packed in a two-dimensional area, which is difficult to accomplish with piezoelectric fabrication, but relatively simple to accomplish with CMUT fabrication.

While CMUT fabrication technology provides an easier, simpler way to fabricate 2-D arrays, a significant barrier to their widespread industrial adoption is the complexity required to actually integrate them with electronics. In standard ultrasound arrays, each element is connected by a separate channel to its required electronics. In a linear array, where 128 or 256 elements are arranged sequentially, this does not pose a problem. However, in a 2-D array, the total number of elements grows quadratically as the array length and width increases, meaning a 2-D array with even a relatively small side width can have thousands or tens of thousands of elements. Wiring each element to a separate channel in a small area becomes prohibitively complex and expensive for large, dense 2-D arrays. Because (in addition to having the ability to image larger areas), larger arrays have better resolution, it is critical to develop novel architectures and solutions which can overcome this problem and promote more widespread adoption of 3-D ultrasound.

1.2 Objectives of Thesis

The overall objective of this thesis is to present a 2-D CMUT array architecture which addresses the high-channel count problem of fully-wired 2-D arrays. The key concept behind this design, which is explored further in chapter 5, is that instead of having individual channels for each element, electrodes are connected orthogonally such that each row and each column of the array is individually addressable. This architecture is referred to fittingly as a 'top-orthogonal-to-bottom-electrode' or 'TOBE' array. With this architecture, the number of channels increases linearly with array size; an array with N x N elements requires only 2N channels rather than N^2 channels, as a fully wired array would. This significantly reduces the complexity and cost of fabricating and packaging larger arrays. The successful implementation and development of this array architecture could significantly increase the number of applications where 3-D ultrasound is viable. It could allow for larger arrays which could potentially image entire organs without the need for operator movement of the array, or it could allow for very small arrays that could fit on the tip of a catheter (where a large number of channels would not be feasible) for use in ultrasonic guidance of keyhole surgeries. Various imaging modalities that would be possible with this array structure could tailor the array for optimal use in many different possible applications, from medical to industrial. The body of this thesis will present the work done on designing, fabricating, characterizing, and simulating these TOBE

ultrasound arrays. In addition, it will also introduce the fundamental physical concepts behind the work so as to demonstrate the thorough understanding and knowledge of CMUT and ultrasound technology that was gained during this process.

1.3 Organization of the Thesis

While the focal point of this thesis is the work devoted to the design and characterization of TOBE CMUTs, it is organized in such a way that a thorough understanding of CMUT and ultrasound technology is demonstrated before it proceeds logically to the methods and results. Chapter 2 begins by providing a background of the ultrasound concepts applied over the course of the work done for this thesis, including fundamental acoustics theory and piezoelectric technology. Chapter 3 follows by providing a more detailed explanation of the principles behind ultrasound imaging specifically. The properties of ultrasound arrays, along with descriptions of different kinds of imaging methods are presented. An emphasis is placed on 3-D imaging, since it is the targeted application for TOBE CMUTs. Chapter 4 shifts the focus to CMUT technology in particular. Fundamental physical principles behind the operation and structure of CMUTs are covered, and descriptions of the various fabrication methods that can be used to make CMUTs are provided and analyzed. Comparisons to piezoelectric technology are presented and discussed. Chapter 5 focuses specifically on the novel aspects of the TOBE CMUT arrays and the work that has been done on them. Further details on both the specific array structure and some of the specific imaging principles are included. Methods used to characterize and simulate the operation of these CMUTs are included along with their results and corresponding discussion. Finally, chapter 6 contains concluding statements, including both a summary of presented results and work, and possibilities for future use.

Chapter 2

Background to Ultrasound

2.1 Introduction to Ultrasound Technology

While the principles of acoustics have been studied since the 18th century, the application of these principles to adapt ultrasound for practical purposes was not widely done until the 20th century, beginning when Paul Langevin developed a method of detecting enemy submarines during World War I [3]. Building on this development, ultrasound technology began being used in a number of therapeutic medical applications, as well as many imaging applications. Understanding the fundamentals of acoustic and ultrasonic theory is an invaluable process that is necessary to develop a better grasp of the many different and useful applications of ultrasound.

2.1.1 Ultrasound Wave Basics

As a basic definition, ultrasound is defined as mechanical vibrations that are above the human range of audible frequencies. While the lower range of ultrasonic frequencies is 20 kHz, it is common for ultrasound applications to use much higher frequencies in the megahertz range. Ultrasound applications with bandwidth even in the gigahertz range have even been demonstrated [4], and with the increasing prevalence of laser based ultrasound generation (photoacoustic effect), higher frequency applications for ultrasound are becoming more common. As mechanical waves, ultrasound waves require a medium to propagate, and this medium can be either a fluid or a solid. Applying the principles of ultrasound propagation in a fluid is especially important for medical imaging, since human tissue generally behaves quite similarly to a fluid. However, for NDT, where ultrasound may be required to propagate through solid materials to characterize their composition or integrity, the principles of solid propagation become equally as important. In both fluids and solids, ultrasound waves propagate as longitudinal waves, where the particle potion in the material is *parallel* to the direction of the displacement of the wave. In solids, shear (or transverse) waves are also supported. These waves are characterized by particle displacement *perpendicular* to the motion of the wave. Unlike longitudinal waves, shear waves do not cause localized changes in the density or the pressure of the medium. Two other wave types are important when specifically relating to ultrasound transducer performance, and the effects of these wave types will be elaborated further in section 5.3. The two waves are known as Lamb waves and Stoneley waves. Lamb waves occur in flat plates with thicknesses of a few wavelengths. They propagate parallel to the direction of the surface of the plate, and cause complicated shear particle displacement in multiple directions. Stoneley waves are high-amplitude waves that travel along the interface between a solid and a fluid. While these kinds of waves have considerably more complex properties that will not be covered in this thesis, their existence is important to note, because they can serve

as a mechanism of inter-element acoustic crosstalk in ultrasound arrays and act as a detriment to ultrasound imaging. They can also be used advantageously, such as in the case of Lamb waves, which are commonly used in NDT [5]. A visualization of the different types of waves is given in figure 2.1.

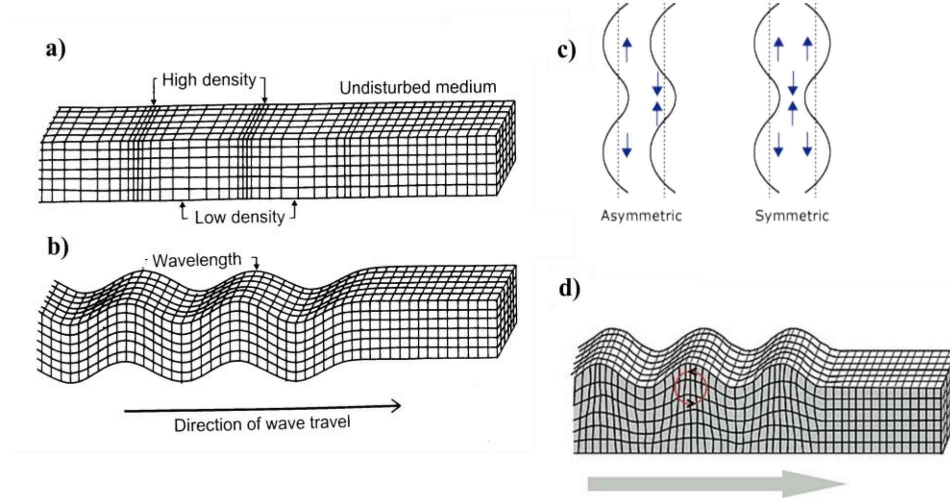


Figure 2.1: Illustration of four types of waves important to ultrasound propagation. a) Longitudinal[1] ; b) Transverse[1] ; c) Lamb[2] ; d) Stoneley[3]

2.1.2 Acoustic Wave Propagation

The characteristics of a wave propagating in a medium are determined by the properties of the medium itself. A wave traveling in a fluid has a propagation speed (c_0) related to the adiabatic compressibility (κ) and the equilibrium density (ρ_0) of the medium and given by the formula:

$$c_0 = 1/\sqrt{\kappa\rho_0} \tag{2.1}$$

In a solid, both longitudinal waves and shear waves exist and have separate propagation speeds that are also dependent on the Young's modulus (E) and Poisson's ratio (σ) of the solid, in addition to its density. The longitudinal wave propagation speed (c_L) and the shear wave propagation speed (c_T) are given respectively by:

$$c_L = \sqrt{\frac{E(1-\sigma)}{\rho_0(1-2\sigma)(1+\sigma)}} \tag{2.2 a}$$

$$c_T = \sqrt{\frac{E}{2\rho_0(1+\sigma)}} \tag{2.2 b}$$

[1] http://www.astro.uwo.ca/~jlandstr/planets/webfigs/earth/slide1.html
[2] http://www.ndt-ed.org/EducationResources/CommunityCollege/Ultrasonics/Physics/modepropagation.htm
[3] http://earthquake.usgs.gov/learn/glossary/?term=Rayleigh%20wave

The propagation speed of any wave is related to its frequency (f) and wavelength (λ) by the equation:

$$c_0 = f\lambda \tag{2.3}$$

and another property of the wave, known as the wave number (k), is given in lossless media by:

$$k = \frac{2\pi}{\lambda} \tag{2.4}$$

Because only longitudinal waves are supported, it is easier to determine the behaviour of a wave in a fluid than in a solid. If an ultrasound wave is propagating in a fluid with velocity $\boldsymbol{v_L}$, then this velocity can also be expressed as the gradient of the velocity potential (ϕ), as given by the equation:

$$\boldsymbol{v_L} = -\nabla\phi \tag{2.5}$$

The linear propagation of this wave in homogenous viscous fluids is governed by the wave equation, which is given by:

$$\kappa\rho_0 \frac{\partial^2\phi}{\partial t^2} = \nabla^2\phi + \kappa\left(\mu_B + \frac{4}{3}\mu\right)\frac{\partial}{\partial t}\left(\nabla^2\phi\right) \tag{2.6}$$

The terms μ and μ_B represent shear and bulk viscosity respectively. These terms are known as the viscosity coefficients and represent the energy dispersal of a wave in a viscous (lossy) medium. If an inviscid fluid is assumed, these terms are zero. For ultrasound imaging, we are usually interested in determining the pressure of an ultrasound wave, and this can be found from the propagation velocity or velocity potential by using the following equations (for an inviscid fluid):

$$p = \rho_0 \frac{\partial\phi}{\partial t} \tag{2.7 a}$$

$$\nabla p = -\rho_0 \frac{\partial \boldsymbol{v_L}}{\partial t} \tag{2.7 b}$$

The velocity potential at any point in space and time is itself determined by the solutions to the wave equation given in (2.6).

2.1.3 Wave Properties at Interfaces

Because ultrasound transducers and the desired medium of ultrasound transmission are separated by a physical boundary, it is useful to understand the principles of wave behaviour at interfaces. An important property of any material is its characteristic acoustic impedance (Z_o), which characterizes its resistance to wave propagation, and is measured in Rayleighs (Rayl). For an inviscid fluid medium, this is defined as:

$$Z_0 = \rho_0 c_0 \tag{2.8}$$

When an ultrasound wave is incident upon an interface between two materials with different properties as shown in figure 2.2, reflection and refraction occur, just as with light. Snell's law defines the angles at which the incident wave is refracted and reflected. The transmitted and reflected wave angles can be found using the following equations respectively:

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{c_{o1}}{c_{o2}} \tag{2.9 a}$$

$$\theta_r = \theta_i \tag{2.9 b}$$

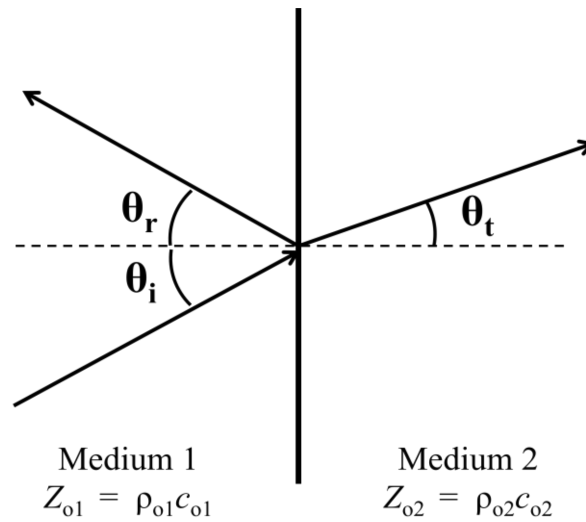


Figure 2.2: Behaviour of a wave at the interface between two media

When a wave is traveling from one medium with a higher propagation speed to another with a lower propagation speed, there are two important angles which exist. The critical angle (θ_c) is the angle at which the transmitted angle is 90 degrees. At this angle, all incident energy is reflected, despite the fact that the transmitted beam has a non-zero amplitude. Alternatively, at the intromission angle (θ_{int}), all of the incident energy is transmitted, and no energy is reflected. The critical and intromission angles can be found respectively from the equations:

$$\theta_c = \sin^{-1}\frac{c_{o1}}{c_{o2}} \tag{2.10 a}$$

$$\sin\theta_{int} = \sqrt{\frac{\left[1-\left(\frac{Z_{o1}}{Z_{o2}}\right)^2\right]}{\left[1-\left(\frac{\rho_{o1}}{\rho_{o2}}\right)^2\right]}} \tag{2.10 b}$$

When transmitting ultrasound from one medium to another, it is desirable to transmit as much energy as possible from the first medium to the second. The amount of energy reflected and transmitted at a boundary is determined by the

intensity reflection coefficient (R_I) and the intensity transmission coefficient (T_I) which are given respectively by:

$$R_I = [\frac{Z_{o2}\cos\theta_i - Z_{o1}\cos\theta_t}{Z_{o2}\cos\theta_i + Z_{o1}\cos\theta_t}]^2 \qquad (2.11\ a)$$

$$T_I = \frac{4Z_{o1}Z_{o2}\cos\theta_i \cos\theta_t}{[Z_{o2}\cos\theta_i + Z_{o1}\cos\theta_t]^2} \qquad (2.11\ b)$$

These constants are related by the equation: $R_I + T_I \frac{\cos\theta_t}{\cos\theta_i} = 1$. For normal incidence, such as when a transducer is placed directly against a transmitting medium, these coefficients reduce to:

$$R_N = [\frac{Z_2 - Z_1}{Z_2 + Z_1}]^2 \qquad (2.12\ a)$$

$$. T_N = \frac{4Z_1Z_2}{[Z_2 + Z_1]^2} \qquad (2.12\ b)$$

When ultrasound is being transmitted from a fluid to a solid, an additional phenomena known as wave mode conversion occurs. As previously mentioned, transverse waves are not supported in fluids but occur readily in solids. As such, a longitudinal wave in a fluid incident on a fluid-solid boundary will transmit both a corresponding longitudinal wave in the solid and a corresponding transverse wave. In such a case, the longitudinal wave and transverse wave each have different velocities and characteristic impedances which are related respectively by the equations:

$$Z_L = \frac{\rho_2 c_L}{\cos\theta_L} \qquad (2.13\ a)$$

$$Z_T = \frac{\rho_2 c_T}{\cos\theta_T} \qquad (2.13\ b)$$

Each of the transmitted waves has a different intensity that can be determined from the separate transmission coefficients for the longitudinal (T_L) and transverse (T_T) waves. These values, in addition to the intensity reflection coefficient, are given by the following equations:

$$R_L = \left[\frac{Z_L \cos^2(2\theta_T) + Z_T \sin^2(2\theta_T) - Z}{Z_L \cos^2(2\theta_T) + Z_T \sin^2(2\theta_T) + Z}\right]^2 \qquad (2.14\ a)$$

$$T_L = \left[\frac{(2\rho/\rho_2) Z_L \cos 2\theta_T}{Z_L \cos^2(2\theta_T) + Z_T \sin^2(2\theta_T) + Z}\right]^2 \left(\frac{\rho_2 \tan\theta_i}{\rho \tan\theta_L}\right) \qquad (2.14\ b)$$

$$T_T = \left[\frac{(-2\rho/\rho_2) Z_T \cos 2\theta_T}{Z_L \cos^2(2\theta_T) + Z_T \sin^2(2\theta_T) + Z}\right]^2 \left(\frac{\rho_2 \tan\theta_i}{\rho \tan\theta_T}\right) \qquad (2.14\ c)$$

The transmission angles for longitudinal (θ_L) and shear (θ_T) components of the wave are related to the incidence angle by the equation:

$$\frac{\sin\theta_i}{c} = \frac{\sin\theta_L}{c_L} = \frac{\sin\theta_T}{c_T} \tag{2.15}$$

The behaviour of ultrasound waves at boundaries between two different materials is critical to understand because it can determine how effective the use of ultrasound in specific situations can be. If most of the energy generated by a transducer is reflected off an interface rather than transmitted, then the transducer's use is limited in that particular situation. This can occur when the characteristic impedance of the transducer material is much different than that of the transmission medium, and this phenomenon is one of the driving factors behind research into CMUTs, as will be explained in greater detail in section 4.2. To maximize the efficiency of a transducer such that most of the energy is transmitted into the desired medium, it is sometimes necessary to include an impedance matching layer between the transducer and the medium, as shown in figure 2.3.

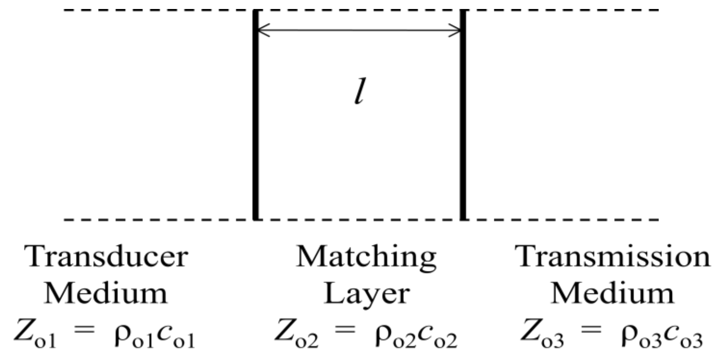


Figure 2.3: Depiction of a matching layer between the ultrasound transducer and the medium

In this case, when normal incidence is assumed, the intensity transmission coefficient from the transducer medium to the transmission medium is given by:

$$T_I = \frac{4Z_{o3}Z_{o1}}{(Z_{o1}+Z_{o3})^2\cos^2\theta_2+(Z_{o2}+Z_{o1}Z_{o3}/Z_{o2})^2\sin^2\theta_2} \tag{2.16}$$

where $\theta_2 = 2\pi l/\lambda_2$, l is the thickness of the matching layer, and λ_2 is the wavelength of the ultrasound wave in the matching layer. It can be readily shown [3] that transmission is maximized when $Z_{o2} = \sqrt{Z_{o1}Z_{o3}}$ and $l = \frac{\lambda_2}{4}$. Because it can be difficult to find a homogeneous medium with the correct matching impedance, the matching layer is usually created by loading a low-impedance material such as epoxy with particles of metal such as tungsten to obtain a 'net' impedance.

2.1.4 Ultrasound Scattering and Absorption

Important to the process of ultrasound imaging are the concepts of ultrasound scattering and absorption. Scattering is the redirection of ultrasound waves due to localized changes in density and compressibility in the propagation medium. Hence scattering would not occur in an ideally pure fluid, such as water. Imaging would not be possible without the concepts of reflection described in the previous section, and reflection is simply a form of scattering where the boundary between two materials represents the change in density and compressibility. Absorption occurs when mechanical energy due to ultrasound is converted into heat or other forms of energy, resulting in a decrease of the total ultrasound energy. As it propagates through a medium, an ultrasound wave will be both scattered and absorbed, and the energy at successive distances throughout the medium will decrease. This decrease in energy due both scattering and absorption is known as attenuation. Every material has an amplitude attenuation coefficient due to both scattering (α_S) and absorption (α_a) and the total amplitude attenuation coefficient is equal to their sum as given by: $\alpha = \alpha_S + \alpha_a$. If a plane wave is normally incident upon a medium (at position 0) with intensity equal to $I(0)$, then its intensity at a point x within the medium can be approximated by:

$$I(x) = I(0)e^{-2\alpha x} \tag{2.17}$$

Determining the exact attenuation coefficient and the amount of scattering and absorption, as well as the overall exact behaviour of a wave in an attenuating medium is significantly more difficult than what has been explained, especially for waves that do not approximate a plane wave. However, one important overarching concept is that the attenuation coefficient is frequency dependent. For a pure fluid with no scattering, attenuation is accounted for by the shear and bulk viscosity coefficients in the wave equation (2.6). A good approximation for the attenuation coefficient in this case is $\alpha = \alpha_0 f^2$ where α_0 is a constant. In biological tissues, where scattering contributes significantly to attenuation, the frequency dependence can be approximated by $\alpha = \alpha_0 f^n$ where n is a constant that varies between 1 and 2 depending on the tissue. The take-away point is that attenuation increases with frequency, and hence high frequency ultrasound cannot penetrate as far into the propagation medium as lower frequencies. Another effect of frequency-dependent attenuation is that when an ultrasound signal is a short pulse containing many frequency components, each component will have a different attenuation (known as dispersion), and the shape of the pulse will change as the signal propagates through the medium, as shown in figure 2.4.

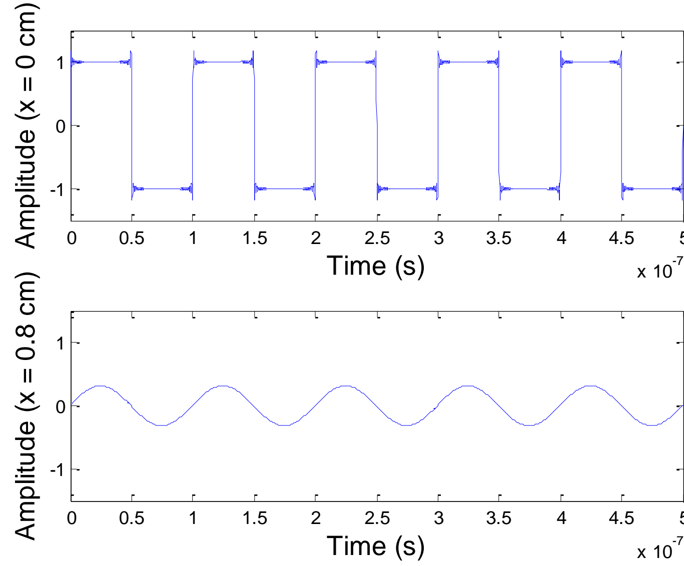


Figure 2.4: Effect of frequency dependent attenuation on signal shape

2.2 Ultrasound Field Profile Analysis

An important part of designing an ultrasound array is using simulations to determine its imaging performance beforehand. The ability of an array to perform imaging is dependent on the ultrasound pressure field profile generated by the transducer, and hence simulating ultrasound image generation requires knowledge of the pressure field profile. This profile is determined by the wave equation given in equation (2.6), and many different methods have been proposed to calculate it.

2.2.1 Analytical Methods for Calculating Ultrasound Fields

The easiest method to determine the field profile from a transducer is to define the transducer as a collection of point sources that each emit a spherical wavefront, and then determine the superposition of those waves at the desired observation point. This method is encapsulated by the Rayleigh integral, which states that at a point **r** on some plane z = z_o, the velocity potential due to a transducer surface on the plane z = 0 is given by the equation:

$$\phi_N(\boldsymbol{r}:t) = \int_{-\infty}^{\infty} v_{no}(\tau) \iint_{S_o} \frac{\xi_o(x_i,y_i)\delta\left(t-\frac{R}{c_o}-\tau\right)}{2\pi R}\, dS_o d\tau \qquad (2.18)$$

where $v_{no}(t)$ represents the time-varying normal velocity of the transducer surface, R is the distance between the observation point and the point on the transducer, and $\xi_o(x,y)$ represents the spatial variation of this velocity across the surface. This equation can be rewritten in the form:

$$\phi_N(\boldsymbol{r}:t) = v_{no}(t) * h_t(\boldsymbol{r}:t) \qquad (2.19)$$

or by expressing the velocity potential in terms in pressure from equation (2.7 a):

$$p(\boldsymbol{r}:t) = \rho_o h_t(\boldsymbol{r}:t) * \frac{\partial v_{no}(t)}{\partial t} \tag{2.20}$$

where $h_t(\boldsymbol{r}:t)$ is the field due to an impulse function at the transducer. This integral is derived from the solutions to the wave equation and assumes that the propagation medium is inviscid and that the velocity at the plane which contains the transducer surface is zero outside of this surface (rigid baffle boundary conditions).

The angular spectrum method is another method by which the field profile can be calculated which takes advantage of efficient 2-D fast Fourier transform (FFT) algorithms to reduce the computation time necessary to calculate the field [6]. It allows the field profile on one plane to be calculated by using the field profile on a parallel plane. When related to the normal velocity at the transducer plane $v_z(x_i, y_i, 0)$, the equation is:

$$\phi(x_o, y_o, z_o) = \iint_{-\infty}^{\infty} v_Z(x_i, y_i, 0)\left[\frac{j}{4\pi^2}\iint_{-\infty}^{\infty}\frac{1}{k_z}\, e^{-j[k_x(x_o-x_i)+k_y(y_o-y_i)+k_z z_o]}dk_x dk_y\right]dx_i dy_i \tag{2.21}$$

where k_x, k_y, and k_z are the spatial frequency components in the *x, y,* and *z* directions respectively. This equation is equivalent to performing a 2-D FFT of the source plane velocity, multiplying it by the transfer function from the source plane to the observation point plane, and taking the inverse 2-D FFT of the resulting product.

2.2.2 Approximations for Calculating Ultrasound Fields

Because it can be difficult to compute the field profile exactly using analytical solutions, it can be helpful to have some approximations which can make the process easier. One such commonly used approximation is the Fresnel approximation. In this approximation, it is assumed that the observation point (x_o, y_o, z_o) is far enough away from the transducer such that the variation in the distance between this point and any point on the transducer surface $(x_i, y_i, 0)$ is negligible compared to the distance itself. If the normal transducer surface velocity is harmonic with angular frequency ω and given by $v_{no}(t) = \xi(x_i, y_i)v_o e^{j\omega t}$, then using the Fresnel approximation, the Rayleigh integral given in equation (2.18) becomes:

$$\phi_N(x_o, y_o, z_o: t) = \frac{v_0 e^{j\omega t}}{2\pi R}\iint_S \xi_0 e^{-j\omega(R/c_0)}dS \tag{2.22}$$

where R can be approximated to be:

$$R \approx z_o(1 + \frac{1}{2z_o^2}[(x_o^2 + y_o^2) - 2(x_o x_i + y_o y_i) + (x_i^2 + y_i^2)] \tag{2.23}$$

If it is further assumed that the observation point is close to the axis passing through the center of the transducer $(x_o = 0, y_o = 0)$, then the last term in equation (2.23) can be neglected and we are left with the Fraunhofer

approximation. By noting the relationship between pressure field and velocity potential given in equation (2.7 a), the pressure phasor at the observation point can be expressed as

$$p(x_o, y_o, z_o : \omega) \approx \frac{j\omega\rho_o v_o e^{j\omega t}}{2\pi R} e^{-jk\left[z_o + \frac{x_0^2 + y_0^2}{2z_o}\right]} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \xi_o(x_i, y_i) e^{-j(k_x x_i + k_y y_i)} dx_i dy_i \quad (2.24)$$

The integral portion of the above equation is simply the 2-D FFT of the function $\xi_o(x_i, y_i)$. As previously mentioned, this function describes the variation of surface normal velocity over the entire plane encompassing the transducer surface. It can be separated into two constituent functions: $\Omega(x_i, y_i)\xi_o(x_i, y_i)$ if necessary, where $\Omega(x_i, y_i)$ is a function that has a value of 1 wherever there exists a transducer surface, and zero everywhere else. When dealing with piston transducers, as many simulations usually do, then $\xi_o(x_i, y_i)$ is simply 1, and $\Omega(x_i, y_i)$ is called the aperture function. Thus, the Fraunhofer approximation allows for the calculation of the pressure field on an observation plane by simply taking the 2-D Fourier transform of the function defining the geometry of the transducer, vastly simplifying the process.

One result of using approximations to determine the pressure field is the division of the desired observation area into different 'approximation zones'. If the normalized distance from the transducer surface is given by $Z_n = z/(a^2/\lambda)$, where a is the radius of the transducer and λ is the ultrasound wavelength, then the *near field* is defined as the region where the normalized distance is less than 1. Outside this area is defined to be the *far-field.* Whereas the Fraunhofer approximation only begins becoming accurate at a normalized distance greater than 2, the Fresnel approximation can be used even in the near-field, with accuracy increasing as the ratio a/λ increases [3]. The Rayleigh distance represents the point at which the pressure amplitude is 50% of the maximum, and is located at a normalized distance of π. Although the derivation will not be included here, a focused transducer allows the Fresnel approximation to reduce to the Fraunhofer approximation when the distance from the transducer is equal to its focal point. Focusing allows far-field approximations to be used in the near-field.

2.3 Piezoelectric Ultrasound Transducer Technology

The focus of this thesis is specifically on capacitive micromachined ultrasound transducers, or CMUTs. However, this technology is still relatively new and commercially unused compared to previously developed technologies that have been in use for decades. To fully appreciate the novelties and advantages of CMUT technology, it is important to understand the basics behind these previous technologies as well. The launching point of modern ultrasound transducer technology occurred in 1880, when piezoelectricity was discovered by the Curie brothers [7]. While piezoelectric technology was not the only transduction mechanism explored for use in ultrasound transducers, it became far and away the most prevalent.

Piezoelectric transducers function based on the piezoelectric effect (and inverse piezoelectric effect). When certain piezoelectric media are subject to pressure they generate a voltage between two conducting surfaces on either side of the crystal. Inversely, when a voltage is applied to the two conducting surfaces, it causes the crystal to deform. Thus a harmonic applied voltage signal could result in a harmonic transmitted pressure signal. There are many of different types of materials and many different fabrication processes that can produce transducers that exhibit this effect. Single crystals such as quartz and Rochelle salt were the first types of materials explored. Following that, piezoelectric ceramics such as lead zirconium titanate (PZT) and barium titanate (BaTi) were developed, and these had the advantage of being able to be fabricated in multiple shapes with any polarization direction. These materials have also been combined with non-piezoelectric materials such as epoxy to form piezoelectric composites, which can have variable acoustic impedances. Even polymers such as polyvinylidene fluoride (PVDF) have exhibited piezoelectric properties.

Actual transducers have been fabricated with a number of different materials depending on the required applications. Desirable qualities for transducers are a high sensitivity (to generate larger signals), a high dielectric constant (to minimize the effects of parasitic capacitances associated with interface electronics and cables), and, perhaps most importantly, a high electromechanical coupling coefficient, which determines how effectively electrical energy is converted into mechanical energy and vice-versa. Single crystals are used for long-life sensors due to their very stable piezoelectric properties (for example, quartz is used extensively in accelerometers, and lithium niobate is used in high temperature acoustic sensors due to its high sensitivity at these temperatures) [8]. They generally have weak piezoelectric properties compared to other materials, however. Out of the piezoelectric composites, the most commonly used (and indeed the most commonly used material overall) is PZT because of its excellent piezoelectric properties (which can be varied by adding dopants during fabrication) [8]. There exist many commercially available varieties of PZT. Outside of piezoceramics, PVDF is popular for high-frequency, large-aperture, high-frequency single element transducers, despite its low electromechanical coupling coefficient [9], as is lithium niobate, which also has the advantage of being lead-free. Piezoceramics are not as suitable for high-frequency applications because these applications require thin transducers, and piezoceramics generally have a large grain size which limits their minimum thickness. Piezoelectric thin films and piezoelectric composites are still the subject of research and are not widely used commercially. Piezoelectric based transducers are typically fabricated by depositing the piezoelectric transducer material in a layer on a substrate, then dicing the layer to form each individual element, after which the cuts are filled with a material such as epoxy to isolate the elements from each other. [10]

Chapter 3

Ultrasound Imaging

3.1 Pulse-echo Pressure Field Analysis

While the fundamental properties of acoustics are important to understand when researching an ultrasound imaging system, the process of imaging itself has evolved into something significantly more complicated than what can be described simply by the acoustic pressure field propagation from a single-element transducer. There are many different types of imaging modalities, that use many different types of transducer architectures, and each one has its own advantages and disadvantages. Considerations must be made in imaging that take into account such factors as frame rate, resolution, signal-to-noise ratio (SNR), and transducer complexity, and often compromises must be made that improve one factor at the cost of the other. The purpose of this thesis is not to cover all these different modalities or transducer architectures, but rather to provide enough information on the principles behind ultrasound imaging such that the specific imaging concepts applied to TOBE CMUTs can be better understood.

An ultrasound image is defined by a set of electrical signals which contains information about the desired medium. The goal of a transducer is to generate these electrical signals from pressure signals incident from some external source. So while an important part of understanding imaging is knowing how pressure fields generated from transducers propagate, as explained in chapter 2, equally as important is understanding how pressure fields incident on transducers can actually be used to generate images, and how certain parameters of a pressure field define the quality of such an image. Electrical signals generated by a transducer are related to the mechanical force (*F(t)*) on the transducer by an electro-mechanical impulse response function. Furthermore, the force on the transducer is a function of the incident pressure. In section 2.2, the pressure at an observation point **r** due to a transducer was given by the Rayleigh integral (2.18), which defines the total pressure as a superposition of pressures due to point source emitters on the surface area of the transducer. Similarly, since force is simply the product of pressure and area, the force on a receiving transducer due to the pressure generated from some point **r** can be calculated by integrating the received pressure signal at each 'point receiver' across the surface of the transducer. For a delta function of pressure, this is given by the equation:

$$h_r(\boldsymbol{r}, t) = 2\int_S \frac{\delta\left(t - \frac{R}{c_o}\right)}{2\pi R}\, dS \qquad (3.1)$$

where, once again, *R* denotes the distance between the point on the transducer and the source of the pressure. This impulse response can then be used to calculate the applied force for any pressure signal by convolving it with the pressure signal as per the equation:

$$F(t|\boldsymbol{r}) = p_o(t) * h_r(\boldsymbol{r}, t) \tag{3.2}$$

The equivalency between the equations for transmitting and receiving is known as the principle of 'reciprocity', and allows us to apply all of the principles derived in section 2.2 for transmission of ultrasound to reception of ultrasound as well. Figure 3.1 illustrates a typical ultrasound imaging scenario, where a transducer at one location transmits (Tx) a pressure signal to image a point scatterer at a vector **r** from that transducer. The scatterer reflects the signal to a receiving (Rx) transducer (it could also be the same transducer, the physics is unchanged) at a vector **r`** from the scatterer.

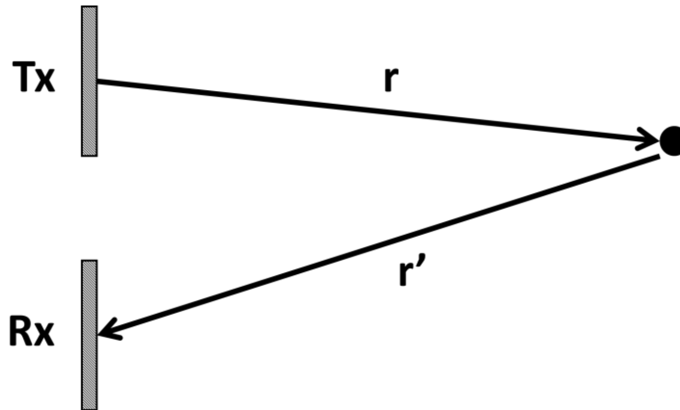


Figure 3.1: A typical transmit-receive ultrasound setup

The pressure at the point scatterer is found using equation (2.20), and, assuming the scattered signal is a constant multiple of the incident pressure, the force on the Rx transducer is therefore given by

$$F(t) = h_t(\boldsymbol{r}: t) * \frac{\partial v_{no}(t)}{\partial t} * h_r(\boldsymbol{r}: t) \tag{3.3}$$

While this equation can give an exact solution for the incident force, the principle of reciprocity also applies to the approximations made in section 2.2.2. This means that for any point in the far-field, by using the Fraunhofer approximation, the radiation pattern seen by the receiver due to that point can be approximated as the 2-D Fourier transform of the receive aperture [**10**]. Thus, by applying the convolution theorem to equation (3.3), it is evident that the full pulse-echo radiation pattern seen at a receive aperture is given by the Fourier transform of the convolution of the transmit aperture function with the receive aperture function. The Fraunhofer approximation would thus also be able to be used for an object at the focal point of a focused receiving transducer as well.

3.2 Linear Array Imaging

3.2.1 Introductory Imaging Concepts

With the ability to predict the pulse-echo pressure field profile of a transducer from its geometry, effective array design for optimal ultrasound imaging becomes significantly easier. In terms of array design, the simplest possible design is a single-element piston transducer. Such a transducer can be used to form what is known as an *A-scan*. This is the most basic form of ultrasound imaging. The signal on the receiving transducer is simply a pressure amplitude vs. time graph characterized by the time taken for a pulse-echo signal to make a round-trip, and an example is shown in figure 3.2 a). To increase the electronic signal strength and imaging range, amplification is usually done on such a receive signal from a transducer. A-scan data can also be visualized in terms of a single line with varying brightness depending on the amplitude of the A-scan. This is known as a B-scan. If B-scans are obtained at multiple positions along the imaging surface, these 'scanlines' can be grouped together to form a B-mode image. This is a 2-D representation of a vertical cross section of the imaging area, where bright spots indicate more reflection, and hence, the presence of interfaces or non homogeneous materials. An example of a B-mode image is shown in Figure 3.2 b). B-mode images can be displayed in real time given a fast enough frame rate for the imaging system. The frame rate is determined by the number of transmit events required to reconstruct an image, and optimizing it is a significant area of research and will be covered in subsequent sections.

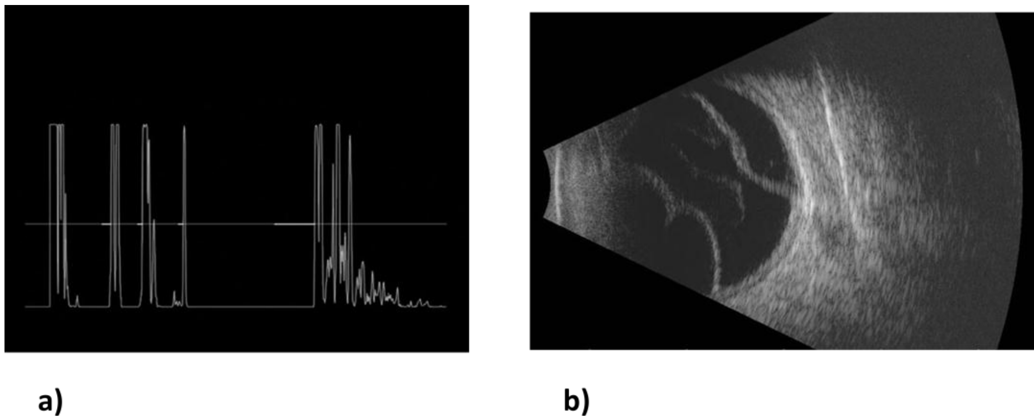


Figure 3.2: a) A-scan b) B-mode image constructed from multiple B-scans[4]

Because of the attenuation effects described in section 2.1, a reflection from a structure deeper within the imaging field would result in weaker pressure signal at the receiving transducer, resulting in a decreased optimal penetration depth. One solution proposed to circumvent these effects was to increase the amplification of the received signal over time [11]. By thus being able to display relatively consistent brightness for the same type of reflections over a greater range of depths in the imaging medium, the quality of B-mode images was significantly

[4] http://www.ellex.com/corp/products/diagnostic-ultrasound/eye-cubed

improved. Almost all medical ultrasound images that are so common and easily recognizable today are B-mode images.

3.2.2 Linear Array Transducer Pressure Field Analysis

While it is possible to create a B-mode image by using a single piston transducer and translating it across the desired cross sectional axis, or tilting it through a specific angle, the effectiveness and quality of this method are limited, which is why for most imaging purposes, single-element transducers have been replaced by transducers with multiple elements, or 'array transducers'. The simplest and most widely used type of array transducer is the linear array. This transducer contains a series of individually controllable elements spaced periodically in a line, as shown in figure 3.3. In all further notation, the *x*-direction refers the direction through the length of the transducer, and is also known as the 'lateral' or sometimes 'azimuthal' direction. The *y*-direction is the direction across the length of the transducer and is known as the 'elevational' direction. Finally, the *z*-direction will be the direction normal to the transducer surface and is also known as the 'axial' direction. To maintain consistent notation, the point (0,0,0) is always located at the center of any transducer.

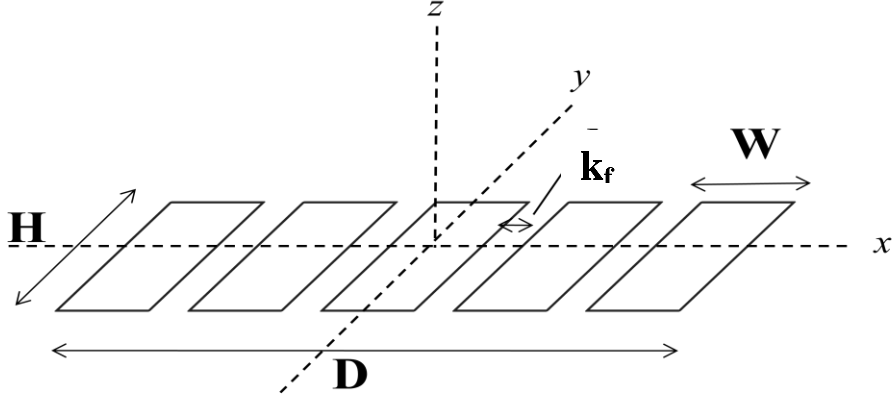


Figure 3.3: A linear array transducer

Using a linear array transducer eliminated the need to mechanically scan a single-element transducer to form a B-mode image. The pressure field profile from this type of array can be found in the same way as the profile from a single element: by taking the 2-D Fourier transform of the aperture function. For an array of *N* rectangular elements that have height *H,* width *W,* spacing between elements k_f (also known as 'kerf'), and total width $D = N\left(W + k_f\right) - k_f$, the aperture function is given by the formula [3]:

$$\Omega(x_i, y_i) = \mathrm{rect}\left(\frac{y_i}{H}\right)\left[\left\{\mathrm{rect}\left(\frac{x_i}{D}\right) \times \sum_{n=-\infty}^{\infty} \delta(x_i - nd)\right\} * \mathrm{rect}\left(\frac{x_1}{W}\right)\right] \tag{3.4}$$

where the *pitch* (d) is equal to the kerf plus the width of an element. The pressure field produced by this aperture on a plane at an observation point (x_o, y_o, z_o) in the far-field or focal plane is then found from the 2-D Fourier transform of this aperture function evaluated at the spatial frequencies $k_x = -\frac{kx_o}{z_o}$ and $k_y = -\frac{ky_o}{z_o}$, (where k is defined by equation (2.4)), which is given by:

$$p(x_o, y_o{:}\,\omega) = \frac{j\omega\rho_o v_o e^{-jkz_o}}{2\pi z_o}\frac{HDW}{d}\text{sinc}\left(\frac{y_o H}{\lambda z_o}\right)\text{sinc}\left(\frac{x_o W}{\lambda z_o}\right) \times \sum_{n=-\infty}^{\infty}\text{sinc}\left[D\left(\frac{x_o}{\lambda z_o} - \frac{n}{d}\right)\right] \tag{3.5}$$

The pressure field as a function of the z_o-normalized lateral distance is shown in figure 3.4. The pressure field in the elevational direction is not shown, but it has a similar profile as the envelope of the lateral field. It should be noted that in all previous and subsequent equations, the apodization, or weighting, of each individual element is assumed to be one for simplicity. It is, however, common to apply apodization functions such as Hanning windows to aperture functions to modify and improve the qualities of generated images.

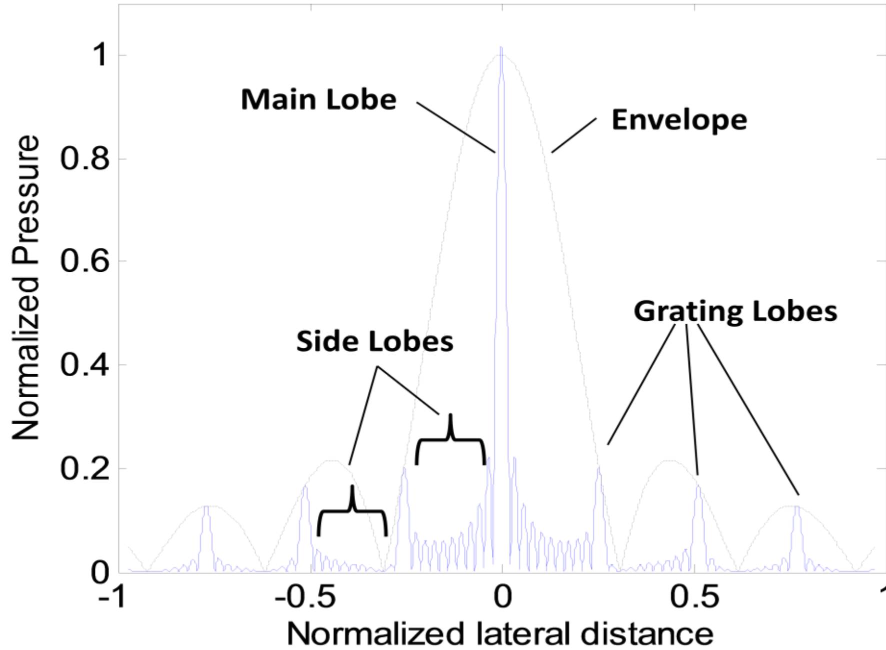


Figure 3.4: Pressure field profile for a linear array transducer

From this pressure profile, it is possible to determine how various properties of the array affect image quality. In figure 3.4, several sections of the pressure profile are labelled. The first is the main lobe, which is the large, central peak of the profile. The grating lobes are the larger prominent peaks symmetrically spaced about the main lobe, while the side lobes are the smaller peaks found in between the main lobes and the grating lobes. The envelope is determined by the function

$\mathrm{sinc}(x_0 W/\lambda z_0)$. To obtain the best possible resolution in an image, it is desirable to have most of the generated pressure confined to a small area, so as to be able to get accurate pulse-echo data from that area without interfering reflections from other areas. To achieve this there are a number of factors which should be considered. Firstly, the main lobe should be as narrow as possible and have a large amplitude compared to the rest of the signal. This correspondingly means that the grating lobes and the side lobes should be as small as possible. In addition to improving resolution, increasing the ratio of energy contained in the main lobe increases signal-to-noise ratio and increases contrast.

Many of the properties of this pressure profile can be determined from the aperture function formula, and hence from the geometry of the aperture itself. Firstly, the width of the main lobe is given by $2\lambda/D$. This indicates that two direct factors that influence image resolution are the size of the array and the wavelength of the ultrasound. Since we want to minimize the width of the main lobe, it is desirable to have a small wavelength (high frequency), and a large array. However, as mentioned in section 2.1.4, high frequency signals suffer more from attenuation, and thus have decreased penetration depths. This necessitates a trade-off between imaging depth and resolution. In figure 3.5 e) and f), the array size has been increased and decreased respectively, resulting in narrower and wider main lobes respectively than in figure 3.5 a). The effects of increasing and decreasing frequency on the main lobe width can be seen in figure 3.5 b) and c) respectively. Because grating lobes are higher amplitude peaks spaced away from the main lobe, they can cause unwanted reflections when the energy from these lobes is reflected by off-axis structures. The grating lobes are spaced apart by a distance of λ/d. Therefore, to have no grating lobes, would require an element spacing of zero (one long element). Since this is not feasible, the grating lobes instead should be spaced as far apart as possible, which would require as small an inter-element distance as possible. It should be noted that decreasing the wavelength, while increasing resolution, also decreases the spacing of the grating lobes, as shown in figure 3.5 b). In figure 3.5 d) the same array as figure 3.5 a) was made sparser by removing elements periodically. As can be seen, the grating lobes distance decrease. A similar effect is observed in figure 3.5 e) where the kerf was increased while keeping the number of elements and their width constant. As mentioned, in the elevational (y) direction, the pressure profile is simply a sinc function. The width of the main lobe is similarly given by the formula: $2\lambda/H$.

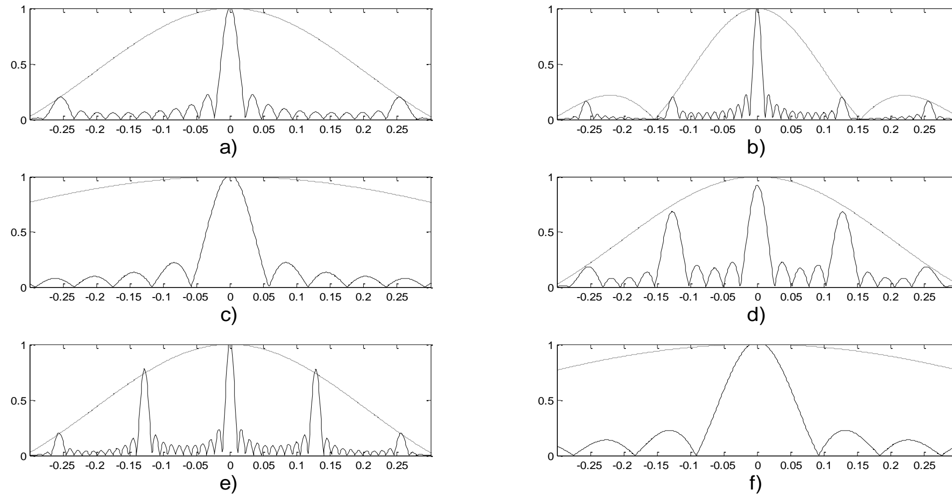


Figure 3.5: Pressure field profiles demonstrating the effect of changing linear array parameters. a) The array has the same properties as the one given in figure 3.4 as a base comparison. b) The frequency has been doubled. c) The frequency has been reduced by 2.5 times. d) The array is the same size, but has fewer elements and an increased kerf. e) The kerf has been increased but the number of elements is the same as in a), resulting in a larger aperture. f) The number of elements has decreased, thus decreasing the aperture size, but all other parameters remained constant

3.2.3 Beamforming with linear arrays

While the pressure field profile discussed above is a useful tool to determine how the properties of an array affect the image quality, its use is limited if we can only make these connections in the far-field region, where the Fraunhofer approximation is valid, since in many cases, imaging occurs within the near-field region. As previously mentioned, focusing allows the Fraunhofer approximation to be used even in the near-field if it is applied to axial distances around the focal point. Focusing an ultrasound transducer thus allows the same conclusions drawn in the previous section (relating pressure field shape and resolution with transducer geometry) to be applied to the pressure field at this focal point. One way of focusing is mechanical focusing; wherein the transducer is either fabricated in a concave manner along some direction, or an acoustic lens is placed on top of a flat transducer to focus the ultrasound by refracting it based on the principles discussed in section 2.1.3. However, the disadvantage with mechanical focusing is that it allows for only one, fixed, focal point. When imaging, it is desirable to be able to focus at multiple points within the imaging area so as to obtain the best possible image quality. This can be done using electronic focusing and steering. Without focusing or steering, all transducer elements fire at the same time, and then the signals received by each element are summed in time as they arrive. This results in an unfocused scanline. Image quality is improved by performing beamforming. A specific type of beamforming used in ultrasound imaging is known as ‘delay-and-sum’ beamforming. Using this method, focusing

and steering the beam can be performed by adding delays both to the signals applied to the transducers to generate a transmit event, and to the received signals before summing them. Arrays that are capable of doing this are known as *phased arrays*.

The first part of obtaining a well focused image is focusing upon ultrasound transmission. By applying a specific delay value to each element and firing the outer elements before the inner elements, the generated pressure fields can arrive at the focal point at the same time and interfere constructively. To focus at a point z_f, the delay value for the n_{th} transducer from the center of an array with N elements and with a pitch of d is given by [12]:

$$\tau_n = \frac{1}{c_o}\left[\sqrt{z_f^2 + (N-1)^2 d^2/4} - \sqrt{z_f^2 + (nd)^2}\right] \tag{3.6}$$

Steering is a method of propagating ultrasound from the transducer at a non-zero angle and allows for focusing at a point off the transducer's central axis. It is an alternative to using rectilinear imaging. In rectilinear imaging, the transducer is larger than the desired imaging area, and focusing is done by actuating only a subsection of the array at a time so that each subsection is focused on-axis. Because steering can focus at points outside of the transducer aperture, it is beneficial to use. To steer the beam at an angle of ϕ_s from the center axis of the transducer, a constant delay of $\Delta\tau = (d/c_o)\sin\phi_s$ is required between successive transducer element firings (beginning at one end of the array). Steering and focusing can be combined to focus the ultrasound at an off-axis point with distance z_f from the center of the array, and the required delay for the n_{th} element from the center of the array is given by:

$$\tau_n = \frac{1}{c_o}\left[\sqrt{z_f^2 + (N-1)^2 d^2/4 + (N-1) z_f p \sin|\phi_s|} - \sqrt{z_f^2 + (nd)^2 - 2n z_f d \sin\phi_s}\right] \tag{3.7}$$

which, for arrays that have a large focal length compared to their aperture size (a large f-number), can be approximated as:

$$\tau_n = \frac{nd\sin\phi_s}{c_o} - \frac{(nd)^2\cos^2\phi_s}{2c_o z_f} - \tau_0 \tag{3.8}$$

where the first term represents delays due to steering, and the second term represents the delay due to focusing. The last term is a constant to ensure non-negative delays.

The next step after focusing during transmission is focusing at a point when receiving the scattered ultrasound signal. Analogous to transmission, the transducer can also be focused upon reception by applying delays to each element, delaying the individual signal from that element a specified amount before summing it with the rest of the signals. In this way, the different distances between the focus point and each element are accounted for, and the signals are summed coherently. With transmission, focusing at different points requires a

separate transmit event for each desired point, each time applying different delays to the individual elements. With reception, however, it is possible to focus at multiple points along the return path of the scattered signal within just one transmit-receive event. This is known as ‘dynamic receive focusing’. Because the signals reflected from scatterers that are farther away from the transducer will take longer to arrive than signals reflected from scatterers that are nearby, if the delays can be changed in between the reception of the signals from the different scatterers, then the receive aperture can be focused at each scattering point as its signal arrives. The result is a much greater depth of field. The required delay is given by a formula similar to that required for steering and focusing upon transmission:

$$\tau_n(z_f, \phi_s) = -\frac{nd \sin \phi_s}{c_o} + \frac{(nd)^2 \cos^2 \phi_s}{2c_o z_f} + \frac{z_f}{c} \tag{3.9}$$

but because a phased array can dynamically focus on reception, the delay can be expressed as a function of time by substituting the equation $z_f = \frac{c_o t}{2}$ into equation (3.9) to obtain:

$$\tau_n(t, \phi_s) = -\frac{nd \sin \phi_s}{c_o} + \frac{(nd)^2 \cos^2 \phi_s}{c_o^2 t} + \frac{t}{2} \tag{3.10}$$

While steering gives the ability to focus ultrasound at off-axis points without using rectilinear imaging, it can also result in degradation of imaging quality in certain situations. Recall that equation 3.5 gave an expression for the pressure field due to a linear array. The effect of steering can be modeled by adding a term which takes this into account, such that:

$$p(x_o, y_o : \omega) = \frac{j\omega\rho_o v_o e^{-jkz_o}}{2\pi z_o} \frac{HDW}{d} \operatorname{sinc}\left(\frac{y_o H}{\lambda z_o}\right) \operatorname{sinc}\left(\frac{x_o W}{\lambda z_o}\right) \times \sum_{n=-\infty}^{\infty} \operatorname{sinc}\left[D\left(\frac{x_o}{\lambda z_o} - \frac{n}{d} - \frac{\Delta\tau}{2\pi d}\right)\right] \tag{3.11}$$

where $\Delta\tau$ represents the phase shift. Steering shifts the main lobe by an angle of $\sin^{-1}(\frac{\Delta\tau}{kd})$, and , by consequence, also shifts the grating lobes by the same amount. Because the pressure profile is enveloped by the function: $\operatorname{sinc}(\frac{x_o W}{\lambda z_o})$, as the grating lobes are shifted closer to an angle of zero, they get larger, and hence diminish the quality of the obtained image. This is demonstrated in figure 3.6. To avoid grating lobes for steering angles of up to 90º, the pitch must be at most $\lambda/2$, where λ is the shortest wavelength of transmitted ultrasound [3]. Although steering angles are usually smaller than 90º and thus allow for larger element pitches, this spacing requirements represents a fundamental problem, particularly in 2-D arrays, that is a motivating factor for the development of TOBE CMUTs.

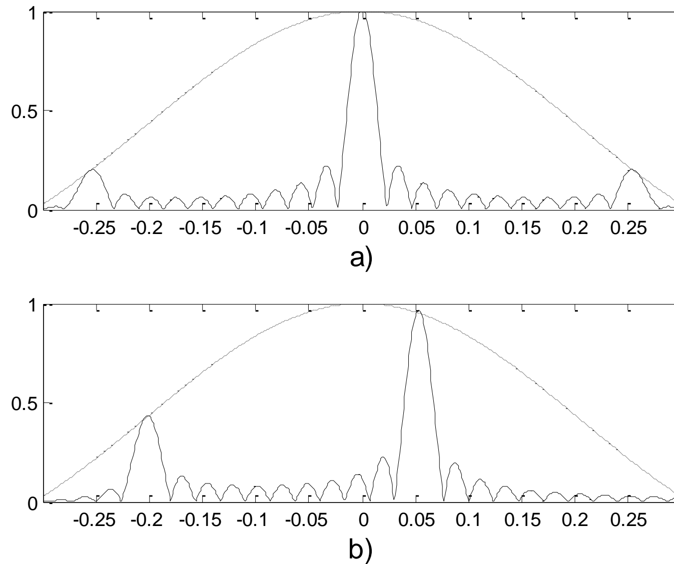


Figure 3.6: Pressure field profile from a linear transducer with a) no steering, and b) 75º steering

The resolution of a linear array determines the minimum size of structures that can be imaged clearly, and it is important to know the resolution limitations of an array so that it can be used in appropriate situations. Each array has a defined resolution in the lateral, axial, and elevational direction. There are a number of definitions for the exact formula of resolution, but an approximate measure for the lateral resolution at the focus point a distance **r** from the transducer center is $\lambda r/D$. While previous discussion has mainly focused on lateral resolution, it is important to have good axial resolution to distinguish smaller structures perpendicular to the transducer surface. This is approximately equal to $c/2B_w$, where B_w is the bandwidth of the transmitted pulse. Unfortunately, increasing the axial resolution by increasing the frequency has the side effect of decreasing the depth of focus, which is approximately given by $6\lambda(R/D)^2$. Elevational resolution is given by the same formula as lateral resolution, only using the aperture size in the elevation direction. For linear arrays, this is generally not an important consideration, but it has significant importance in 3-D arrays.

3.3 Three-dimensional Ultrasound Imaging

While 2-D ultrasound has been used prolifically and successfully in many different applications for many years, the desire to provide even better and more informative images beyond the limitations of 2-D ultrasound has driven research towards the development and improvement of 3-D ultrasound. 3-D ultrasound was demonstrated in clinical applications as early as 1961 [13], and since then has especially seen increased use in medical applications such as imaging of soft tissues such as kidneys, livers, breasts, and foetuses [14], as well as bracytherapy and echocardiography [15]. It has also seen growth in NDT applications [16]. The advantages of 3-D ultrasound lie mainly in its ability to directly provide

volumetric information about the imaging area. In medicine, the study and analysis of 3-D anatomy with 2-D ultrasound imaging relies on the mental transformation by a diagnostician of a series of 2-D images into a 3-D visualization. This subjective analysis can lead to errors in estimating quantities such as organ volume and possible inaccuracies in providing proper diagnoses. Moreover, it can be difficult to localize or reproduce a specific image plane using 2-D ultrasound, making repeatable measurements more difficult. 3-D ultrasound removes much of the subjectivity in imaging by using a computer to process the various 2-D planes and form a 3-D image. Certain methods of 3-D imaging can ensure much greater repeatability in locating specific image planes by almost entirely eliminating the need for mechanical input by an operator. One type of imaging which is made considerably easier with 3-D ultrasound is *C-mode* imaging, where rather than being perpendicular to the ultrasound transducer surface, the image plane is parallel to the transducer surface (i.e. it provides no depth information). Initially, progress in 3-D imaging was slow, because the computational power needed to reconstruct 3-D images was very large, but improvements in both computer technology and reconstruction techniques have allowed for much more flexibility in this regard [15]. For these reasons, significant study has been done on a variety of different 3-D imaging techniques in an effort to promote greater ease and objectivity in imaging.

3.3.1 Three-dimensional Imaging using Linear Arrays

The development of 3-D ultrasound imaging using one dimensional (1-D) or linear arrays began accelerating in the 1990's [3]. In this method of imaging, a linear array obtains a series of 2-D image planes as it is physically scanned, and these images are then pieced together by a computer to form a 3-D image. The scanning can be done either automatically (using a motorized device to move the linear array) or by the hand of an operator (freehand acquisition). While freehand is simpler and allows for more flexibility in transducer positioning, measuring the relative position of the linear array becomes more difficult. The advantage of mechanical scanning is that a specific number of images can be obtained at set intervals so that the required volume is sampled uniformly [3]. Also, the relative position of each 2-D slice is precisely known. The three main types of scanning are linear scanning, where the linear array is translated linearly across the desired area, tilt scanning, where the transducer is tilted about an axis parallel to the transducer face, and rotational scanning, where the linear array is rotated about an axis perpendicular to the transducer face [15]. It should be noted that because the images are obtained with linear arrays, it is still not possible during transmit events to focus in the elevational direction, unless an acoustic lens is used to provide a fixed focus.

3.3.2 Two-dimensional Planar Arrays

More recently, 3-D ultrasound systems have been constructed using two-dimensional plane arrays. In these arrays, elements are spaced in a 2-D arrangement on a plane, as shown in figure 3.7, rather than simply a line, as with linear arrays. The basic 2-D array contains equally spaced individually

controllable elements arranged in a square pattern. If every element is individually controllable, this is known as a *fully-populated* or *fully-wired* array. The principles of operation for such arrays are easy to derive by extending the concepts discussed in the previous section for linear arrays to two directions. The pressure field can still be found by taking the 2-D Fourier transform of the aperture function, and the relationships between pressure field and imaging performance still hold in the elevational direction. The only difference is that the elevational imaging performance factors such as resolution and presence of grating lobes depend on the array properties in the elevational direction. For example, elevational resolution depends on the size of the 2-D array in the y-direction (D_y) (among other factors) and grating lobe spacing depends on the kerf in the y-direction (k_{fy}). Also analogous to linear arrays is the method of focusing and steering the beam. By extending equations (3.6) and (3.7) to two dimensions, delays can be applied to each element so that ultrasound can be steered through a polar angle

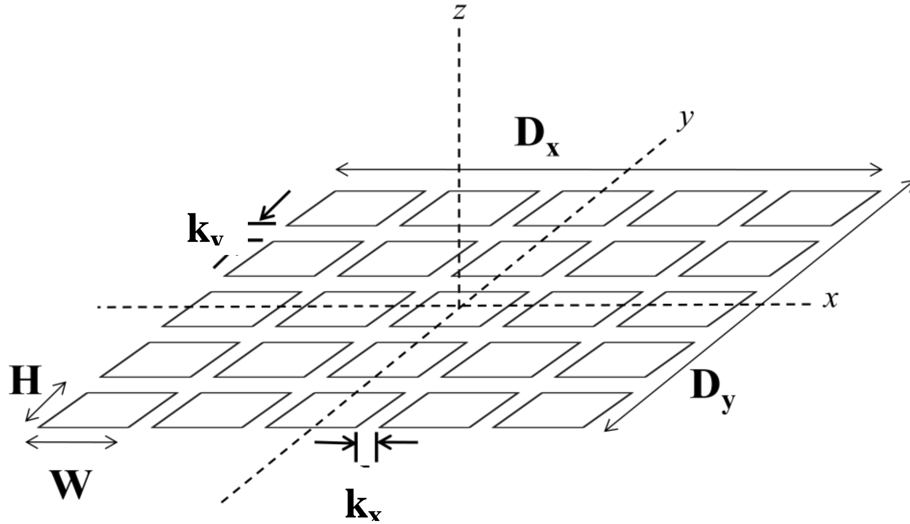


Figure 3.7: A 2-D plane array transducer

(θ_s), and an azimuth angle φ_s, and focused at any point in the imaging volume. If the pitch in x- and y-directions is given by d_x and d_y respectively, then the steering delays between successive elements would be given by:

$$\Delta\tau_x = (d_X/c_o)\sin\theta_s\cos\phi_s \quad (3.12\text{ a})$$

$$\Delta\tau_y = \left(d_y/c_o\right)\sin\theta_s\cos\phi_s \quad (3.12\text{ b})$$

whereas the on-axis focusing delay for the i_{th} and j_{th} elements from the center in the x and y direction respectively would be:

$$\tau_{ij} = \frac{1}{c_o}[\sqrt{z_f^2 + (N_x-1)^2 d_x^2/4 + \left(N_y-1\right)^2 d_y^2/4} - \sqrt{z_f^2 + (id_x)^2 + \left(jd_y\right)^2} \quad (3.13)$$

The combined delay for steering and focusing at any (x_f, y_f, z_f) point is given by [17]:

$$\Delta\tau_{ij} = \frac{R}{c}\left\{1 - \sqrt{(u_o - id_x/R)^2 + \left(v_o - jd_y/R\right)^2 + \cos^2\theta_s}\right\} \quad (3.14)$$

where $R = \sqrt{x_f^2 + y_f^2 + z_f^2}$, $u_o = \sin\theta_s \cos\phi_s$, and $v_o = \sin\theta_s \sin\ \phi_s$. In a fully populated 2-D array which employs focusing and steering (known as a 2-D phased array), ultrasound pulses diverge from the array, sweeping in a pyramidal shape and allowing beam steering and focusing in both elevational and azimuthal directions, so as to display multiple planes from the volume in real time [14]. This removes the need for mechanical scanning and allows the transducer to remain in one position. Alternatively, rectilinear imaging can also be applied to 2-D arrays to avoid steering by using a subset of the full array to do on-axis focusing at each desired point on the x-y plane (using a separate subset for each focal point).

While high-quality images have been produced using physically scanned linear arrays, the method has not seen routine clinical use because of its lack of ability to effectively deliver 3-D images in real time [18]. Not only does the slow scanning speed prevent the acquisition of multiple image planes in a short period of time, but it also requires that a patient remain immobilized to ensure that the structure which is being imaged does not move or change between consecutive acquisitions. These problems can become especially apparent when imaging an organ such as the heart. In this regard, two-dimensional arrays can provide a significant improvement and are more conducive towards real-time 3-D imaging. However, fully-wired 2-D arrays have problems of their own that have prevented their widespread adoption. The first real-time 3-D imaging system was developed by Smith *et al* in 1991 at Duke University [19]. This array was composed of 20 x 20 elements and used the pyramidal 2-D beam steering method described above. However, the constraints of their imaging system allowed for only 32 transmit channels and 32 receive channels. Because the 2-D array had 400 elements in total, only a small number of elements in the array were actually able to be used to produce images. This constraint in their array operation illustrates one of the main disadvantages with 2-D arrays and is the motivation behind research done to develop novel architectures and imaging schemes that can overcome this limitation.

As previously discussed in section 3.2, grating lobes can arise in the pressure profiles of multi-element ultrasound arrays, but can be avoided if the inter-element spacing is less than half the shortest wavelength. In linear arrays, this requirement applies only in the lateral direction, meaning that there is no limit to element size in the elevational direction [20]. For 2-D arrays, however, this constraint applies in the elevational direction as well. In addition, since resolution increases with the size of the aperture, it is more advantageous to have a larger aperture. The combination of these factors means that the ability to obtain high quality images requires a 2-D array that contains a large number of densely

packed, small elements. As an example, while a 64-element linear array would be trivial to implement and fabricate, a 2-D array with a 64-element side length would require 4096 elements, and hence 4096 channels for individual element control – almost two orders of magnitude greater than in the linear array. For arrays with a side length of 128 or 256 elements, this problem is even further compounded. This requirement for a large number of elements and channels creates significant complexity, and hence high cost, in the fabrication and wiring of both the transducer and the interfacing electronics. It is especially difficult to fabricate these arrays using conventional piezoelectric substrates, such as PZT or PMN-PT [21], and in addition, the small element size increases the impedance of an element significantly, making impedance matching of 2-D array elements an additional challenge [22]. Often, front-end integrated circuits are required to preserve signal integrity from these small elements [23].

Another factor which limits the potential size of a 2-D array is the imaging speed. A typical B-mode scan (2-D image) contains between 100 and 250 scan lines per image, which results in frame rates of up to 50 frames per second [24]. To maintain comparable image quality in both the lateral and elevational direction using traditional scanning methods, a similar number of scan lines should be used. However, even with only 70 scan lines in either direction, a volumetric image will contain almost 5000 lines, which, if scanned one-by-one, will be acquired in a period of time too slow for dynamic visualization of the desired volume [24]. For such applications as anatomical imaging, this can be a large enough deterrent to prevent the adoption of 3-D imaging. It is for these reasons that research is being done on numerous ways to simplify the complexity of two-dimensional arrays in such a way that can both increase data acquisition speed and reduce the required number of channels. However, achieving these goals does not necessarily come without a cost. There is always a balance that needs to be struck between array complexity and the quality of the produced image.

3.3.3 Alternative 3-D Imaging Solutions

To overcome the aforementioned problems of imaging speed and system complexity, significant effort has been put into designing alternative array architectures which require fewer elements to perform 3-D imaging. One initial concept was to design 'intermediate arrays' between 1-D and 2-D arrays. The motivation behind these arrays was to add some form of focus control in the elevational plane while only marginally increasing the complexity. Arrays have been designed that have been called 1.25-D arrays, 1.5-D arrays, and 1.75-D arrays [25]. Examples of the first two are shown in figure 3.8. The concept behind 1.25D arrays is the ability to vary the aperture size without being able to perform dynamic focusing. One example provided in [26] divides a linear transducer into three rows, the outer two of which are connected in parallel and have a total area equal to the middle row. By using a multiplexor, different transmit delays can be set for each combination of active imaging rows such that lateral focusing can be individually provided in both the far-field and in the near-field. While imaging ability is improved over a 1-D array, a lens is still the only method of focusing in

the elevational direction, and so only a single beamforming channel is used for all the elements.

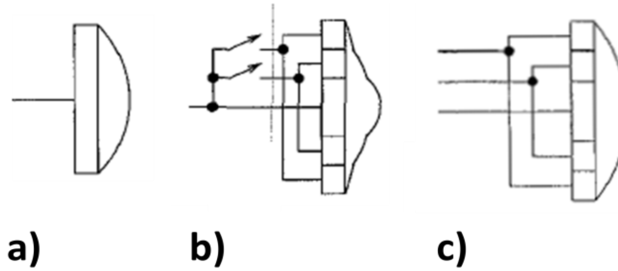


Figure 3.8: Beamforming arrangements for a) Linear array transducer; b) 1.25-D transducer; c) 1.5-D Transducer [26] © 1997 IEEE

The structure of a 1.5-D array is the same as that of a 1.25-D array in that a linear transducer is divided into multiple rows. However, the ability to perform dynamic elevation focusing is added to the array. This means that each group of rows requires its own beamforming channel. In a 1.5-D array, the ability to perform electronic focusing is limited to being symmetric along the elevation axis, meaning that beam steering is not possible. Overall, 1.5-D arrays provide much better contrast resolution than 1-D or 1.25-D arrays, but can require two or three times the number of beamformer channels as a 1-D probe to maintain lateral performance [26]. 1.75-D arrays are identical to 1.5-D arrays; however, the independent control of all elements in the array regardless of symmetry allow for a small amount of beam steering (to within a few degrees). Imaging quality is further increased, but at the cost that each transducer element requires a separate beamforming channel. Continuing to increase the number of rows and decrease the element size in the elevational direction eventually leads to a transducer which does not require a fixed lens and can provide large beam steering angles and dynamic focusing in the elevational direction – that is – a true 2-D array.

3.3.4 Sparse Arrays

The motivation behind sparse arrays was to design a fully 2-D array that did not have the $\lambda/2$ spacing requirement present in conventional phased arrays. Turnbull et al discovered in [17] that the presence of grating lobes was dependent on the arrangement of the elements – specifically on their periodicity. By removing certain elements from a 2-D array and creating a sparsely populated array (sparse array), it was proposed that grating lobes could be mitigated by eliminating the element periodicity, thus circumventing the spacing requirement. In this way, sparse arrays provided the potential for grating lobe-free imaging while reducing the number of channels, and hence, the system complexity. Since their inception, significant research has been done on the best way to distribute the elements in a sparse array. G.R. Lockwood showed that the fundamental premise behind the design of any sparse array is the effective aperture concept [27]. The element distribution in a sparse array must be related to the desired pulse-echo radiation pattern of the array, and the effective aperture concept provides a mathematical solution to the design of a sparse array. The effective

aperture or ‘co-array’ is the equivalent receive aperture that would produce an identical two-way radiation pattern if the transmit aperture were a point source. As previously mentioned, the two-way radiation pattern of an array can be found by taking the Fourier transform of the convolution of the transmit and receive apertures. The effective aperture is simply equal to this convolution. As such, by using the effective aperture concept, the desired transmit and receive arrays can be found by determining the effective aperture required to produce a desired radiation pattern, and then finding the arrangement of elements that best approximates this effective aperture. This pulse-echo radiation pattern should have as low a sidelobe level as possible for a specified main lobe width for all angles and depths of interest [28]. The idea is to obtain an effective aperture which closely mimics the radiation pattern of a conventional 2-D phased array. The specific choice of transmit and receive aperture functions is not trivial, and significant research has been done on algorithms and numerical methods to find the best aperture functions. The goal of this thesis is not to go into detail on the different kinds of sparse arrays, but a brief overview of some basic types of sparse arrays and how their design affects image quality will be covered. In general, sparse arrays are divided into periodic, optimized, and random arrays [28]. Examples are given in figure 3.9.

Periodic sparse arrays retain the periodicity of conventional 2-D arrays to some extent. The transmit and receive elements are spaced periodically, but each of the apertures has a different periodic element spacing. Because they are periodic, the transmit and receive arrays must be arranged such that the transmit and receive grating lobe contributions to the two-way radiation pattern cancel out, thus suppressing the grating lobes. The Vernier interpolation method for 2-D sparse arrays proposed by Lockwood [29] is one of the more common examples of sparse periodic arrays, and has a reputation for having good imaging quality due to its lower integrated sidelobe ratio (ISLR - ratio of energy in sidelobes to energy in main lobe) in comparison to other sparse arrays. Vernier arrays are often used in tandem with fully populated 2-D arrays as a reference for array performance comparisons [30]. Optimized arrays use a specific arrangement of non-periodic elements to create the effective aperture function. Optimized arrays require advanced algorithms to determine their layouts, and because of this, can be very flexible, but also difficult and time-consuming to design given the scope of the amount of possibilities available in removing elements from a large array and determining which selection provides the best pulse-echo radiation pattern. Finally, random arrays remove the periodicity in 2-D arrays by removing a random selection of elements from the array. They can further reduce some of the complexity in optimized array designs, and similar to the latter method, there have been many different methods proposed to create random arrays in an attempt to best replicate the results obtained from sparse periodic arrays.

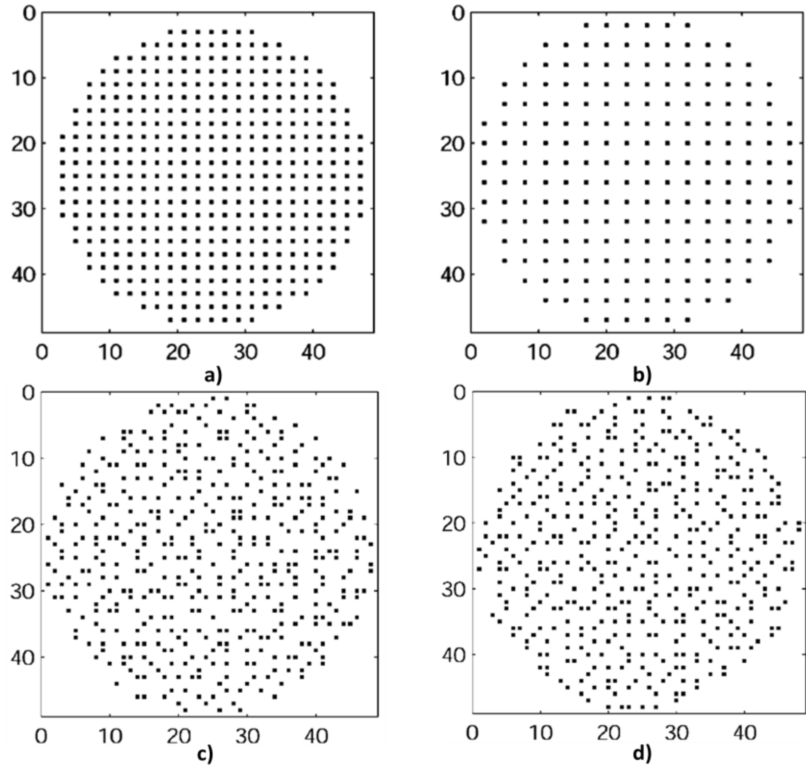


Figure 3.9: Examples of sparse 2-D array designs: a)Vernier (periodic) sparse transmit array; b)Vernier (periodic) sparse receive array; c)Random sparse transmit array; d)Random sparse receive array [30] © 1999 IEEE

Overall, while there exists too much literature to provide a complete description of work done on sparse arrays, conclusions can be drawn that give a good idea of the performance of sparse arrays in general. An important conclusion presented by Turnbull et al. [17] and supported by further literature is that it is possible to remove elements from a fully populated 2-D array without significantly affecting the width of the main lobe. However, the removal of elements results in an increase in sidelobe levels. This is indeed supported by the simulation shown in figure 3.5 d). The width of the mainlobe is nearly identical to that of figure 3.5 a), but the sidelobe levels have increased. It should be noted that in figure 3.5 d), elements were removed periodically to increase the inter-element spacing, which thus resulted in increased grating lobes, and would likely not be a suitable sparse array design on its own. The various types of sparse arrays differ in the location of energy concentration in their sidelobes. For example, the Vernier arrays provide a good approximation to fully-wired arrays, but have high peak sidelobe levels, which can lead to poor low contrast imaging and clutter [31]. Various random arrays such as those proposed by Turnbull et al [17] or Weber et al [32] have smaller peak sidelobes, but exhibit an increased overall sidelobe energy. In general, adding more elements to a sparse array will improve its performance and bring it closer to that of a fully-wired array. For example, Austeng et al. developed an improved periodic arrangement that approximated the response of a fully-wired 2-D array even better than a Vernier array, but only

reduced the number of elements by a factor of 2 compared to a fully-wired array [30]. By comparison, Yen et al simulated a Vernier array with approximately one third as many total elements used as the one simulated by Austeng et al, and found that it performed more poorly in all regards than a random array with approximately equal number of elements [31]. Therefore, the goal of sparse array design is to determine the best compromise between sidelobe energy distribution and number of elements. Overall, it can be concluded that sparse arrays could potentially provide good quality imaging, especially in high contrast areas where the higher sidelobe levels would not be as detrimental. It should also be noted that much of the work done on sparse arrays has involved simulations only. More work will need to be done on verifying how they work in practice before their true value is determined. In most of the simulations performed in published literature, the number of elements was usually no less than 20 times smaller than the number of elements in a fully-wired array of the same size. For smaller arrays, this decrease could be substantial enough to provide a significant benefit. However, because the number of elements in a fully-wired array increases quadratically, the decrease would become less significant for very large arrays. This is a significant motivation to find array architectures that even further decrease the number of channels required, such as the TOBE design.

3.4 Synthetic Aperture Imaging

While the goal of sparse array design was to reduce array complexity by reducing the number of elements, synthetic aperture imaging (SAI) was developed as a method to reduce complexity by reducing the number of elements firing at once, thus reducing the number of total electronic channels required. In most synthetic aperture designs, the frame rate is dependent on the number of array elements used rather than the number of scanlines, and hence another goal of this imaging method was to significantly increase the imaging frame rate. While SAI has mostly been applied to linear arrays, the concepts behind synthetic aperture imaging are relevant to the imaging schemes that are used with the TOBE CMUTs. This section will begin by describing the principles behind SAI for linear arrays, after which they can be extended to 2-D arrays.

3.4.1 Synthetic Aperture Imaging Basics

The most basic type of synthetic aperture imaging, known as the ‘monostatic’ approach or as ‘generic synthetic aperture imaging’, uses only one transducer as both a transmitter and receiver [33] [34]. In the case of a linear array, only one element in the array is used at a time. Rather than display the obtained A or B-scan line data from each emission/reception, the receive data from all emissions is saved, and then all the signals are processed at once. Multiple methods have been proposed to reconstruct an image from synthetic aperture imaging; one common method involves delay-and-sum beamforming in the time domain, similar to conventional imaging. After all transmit-receive events have been performed, each element n has an associated set of receive data $r_n(t)$, which can be time-matched-filtered to improve its quality [35]. Focusing

can then be done at any point within the imaging field retroactively. At such a point (x_f, z_f), the value of the total received signal can be found by the equation:

$$f(x_f, z_f) = \sum_n r_n \left(\frac{2\sqrt{z_f^2 + (x_f - x_n)^2}}{c} \right) \tag{3.14}$$

where x_n is the distance of transducer element *n* from the center of the array. This method essentially sums all the components of all the received signals that occur at the time it would take for the transmitted signal to make a round-trip from the desired focus point. If any of the digitally sampled receive signals do not have a value for this specific time, then it must be interpolated [35]. A visualization of this method is shown in figure 3.10. It should be noted that the delay for SAI is a round-trip delay rather than a one-way delay as in a phased array imaging system. This is because for a phased array, the ultrasound at the focal point is close to a plane wave, whereas each element in the SAI system can be approximated by a point source, which generates a spherical wave. While performance of this method has been shown to be poorer than a phased array [35], only *N* transmit events are required to obtain a full set of image data, where *N* is the number of elements. For a phased array, the number of transmit events is equal to the number of scanlines, which generally should be greater than the number of elements to obtain good resolution.

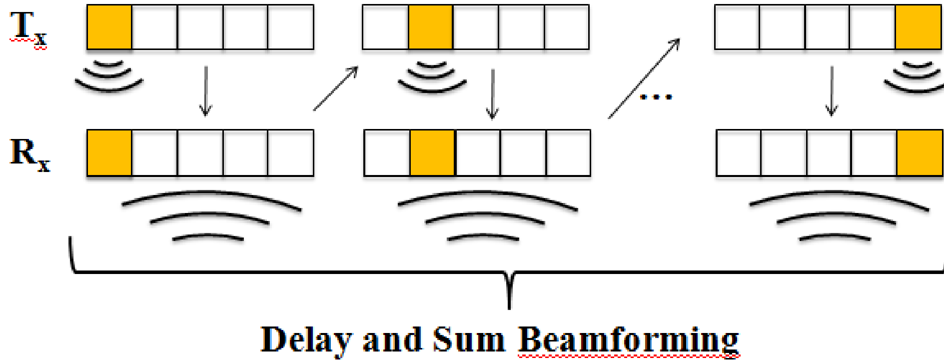


Figure 3.10: Visualization of the monostatic synthetic aperture imaging approach

A second SAI method that is commonly used is known as synthetic transmit aperture imaging (STAI). In this method, a single element is used for transmitting, but all elements are used upon receiving. In this way, the signals from all elements due to one transmit event can be beamformed (delayed and summed) immediately to form a low resolution image. This is done for every element in the array, and then all the low resolution images are summed to produce one high resolution image. For a single transmit event at an element *i*, the low resolution scanline *l* is given by:

$$L_{il}(t) = \sum_{j=1}^{N} r_{ij}(t - \tau_{ij}) \tag{3.15}$$

where $r_{ij}(t)$ is the received signal at element j due to transmission from element i. The delay value for a single focus is given by [35]:

$$\tau_{ij} = \frac{1}{c}\left(|x_o - x_i| + |x_o - x_j|\right) - \frac{2}{c}|x_o - x_c| \quad (3.16)$$

where x_i and x_j are the positions of the transmitting and receiving elements respectively. The high resolution scanline can then be formed simply by directly adding all of the low resolution scanlines together. $H_{il}(t) = \sum_{i=1}^{N} L_{il}(t)$. An advantage of receiving using all the elements at once is that dynamic focusing can be applied upon receive. In this case, the delay τ_{ij} would be a function of time. A visualization of this method is shown in figure 3.11.

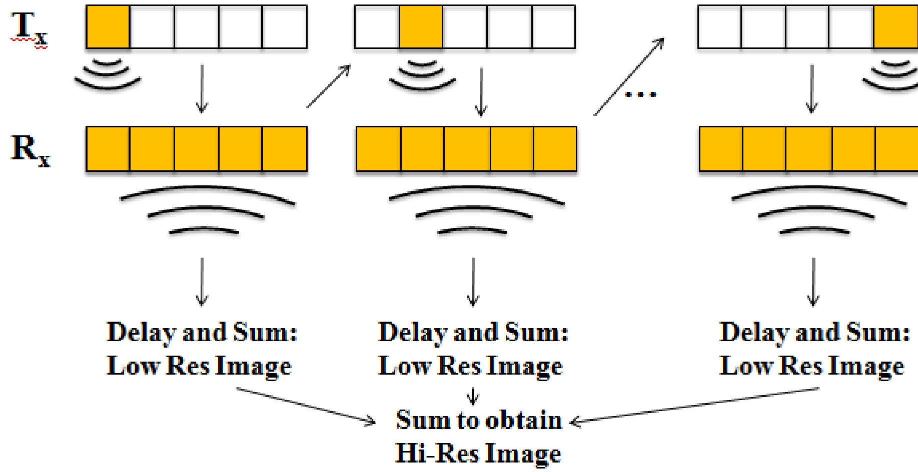


Figure 3.11: Visualization of the synthetic transmit aperture imaging method

Yet another type of synthetic aperture imaging is synthetic receive aperture imaging (SRAI). Because receive electronics are more complicated than transmit electronics, requiring filters, amplifiers, and ADCs among others, it is beneficial to specifically reduce the number of receive channels in a system, which is the goal of SRAI. In SRAI, all of the transducer elements are used during transmit, and only one element is used during receive. The transmit events are focused along a scanline, and for each scanline, there is one focused transmission for each element in the array used for receive. Thus the total number of transmit events is equal to $N \times S$ where S is the number of scanlines and N is the number of elements. While the electronics of this system are simpler, and the obtainable resolution is equal to that of STAI, it requires significantly more transmit events [35]. There have been many different variations of synthetic aperture imaging researched to determine how to obtain the best possible reconstruction of the image. Different combinations of elements on transmit and receive have been proposed so as to offer the best possible image quality. SAI has been combined with sparse linear arrays to further attempt to simplify imaging. Overall, the concept behind all of the different methods is the same: various subsections of the

total array are active upon transmit and/or receive for each transmit event, thus 'synthesizing' the full aperture in multiple steps. Delays are applied retroactively to focus the image at every point.

3.4.2 Synthetic Aperture Imaging for 2-D arrays

Most of the research done on synthetic aperture imaging has focused on 1-D arrays. There is significantly less literature on the topic of SAI applied to 2-D arrays. Much of the research into 3-D imaging using the SAI technique has been done on mechanically scanned 1-D arrays [36] [18]. However, Daher and Yen [37]and Nikolov and Jensen [24] have shown through simulations that the theory behind this imaging method can be carried over to 2-D arrays as well, and the beamforming concepts are the same as those in linear arrays, but extended to three dimensions. The specific methods of beamforming differ for each variation of 3-D SAI, and a further discussion of the beamforming methods used in our TOBE arrays will be provided in section 5.2. One potential problem that arises when using only a few elements to transmit signals is that the weak signal can result in a low signal-to-noise ratio. Using fewer elements when receiving can diminish sensitivity. Thus, many variations on SAI use a larger number of elements to transmit and receive. This concept will also be further discussed in section 5.2 as it applies to TOBE arrays. Rather than stepping an individual element across the entire array, Nikolov and Jensen suggest using a de-focused sub-aperture of the overall 2-D array in an effort to increase pulse energy, using 'de-focusing' delays during transit to approximate a spherical wavefront propagating from a 'virtual' element. In their work, the beamforming method is identical to that described in STAI, except that virtual transmit elements are used, and the delays incorporate three dimensions. In [37] Daher and Yen introduce a row column addressable 2-D array architecture with piezoelectric technology similar to the TOBE arrays that will be discussed. The transmit signals are routed through bottom electrode columns and the top electrodes are connected in orthogonal rows to ground through a switch. They propose three different imaging schemes using synthetic aperture imaging that prioritize either volume acquisition time (given by the number of transmit events), resolution and contrast (given by the -6dB and -20dB beam width), or sensitivity (given by the peak amplitude of the received signal). One of their imaging schemes fires each individual element in a row while receiving along all the elements along this same row, similar to STAI for a linear array, and then repeats the process for each row in the 2-D array. This method optimizes image quality at the expense of a slower imaging time and low sensitivity. Another opts to synthesize the receive aperture row-by-row after firing all the elements in the 2-D array simultaneously, resulting in fast data acquisition at the expense of poor image quality due to lack of transmit focus. This is similar to SRAI where each element is comparable to a single row in the 2-D array. The third integrates synthetic aperture imaging with a sparse array configuration, in an attempt to both reduce system complexity and increase frame rate. As with linear arrays, there are many variations on synthetic aperture imaging that can be applied to different 2-D array architectures.

Chapter 4

Capacitive Micromachined Ultrasound Transducers

4.1 Introduction to CMUTs

In section 2.3, ultrasound piezoelectric technology for ultrasound transducers was introduced. As previously mentioned, these transducers are fabricated from a certain type of material (called a piezoelectric material) that responds to an applied electric field by constricting or expanding, and that generates an electric field when it is subjected to external pressure. Capacitive micromachined ultrasound transducers (CMUTs) operate using a different physical principle – that of electrostatic attraction. The concept of electrostatic transducers is not a novel one – the idea was developed at the same time as early piezoelectric transducers. However, in the past, capacitive transducers were not popular due to their inability to generate the high electric fields needed for efficient ultrasonic wave generation – up to one million volts per centimetre [38]. However, within the past two decades, the technology that allowed electrostatic transducers to achieve such fields became increasingly more prevalent. Advances in silicon-based microfabrication techniques led to the development of microelectromechanical systems (MEMS), and it was this MEMS-based technology that formed the basis behind CMUT fabrication. Not only were these techniques able to create the small gaps between the two capacitor electrodes necessary to achieve the required high electric fields, but they allowed for transducers with improved reliability and performance to be fabricated. Although the first CMUT transducers were not able to compete on a performance basis with piezoelectric transducers, it became clear that the advantages and flexibility that CMUTs potentially offered could solve some significant problems inherent in piezoelectric transducers.

4.1.1 CMUT Structure

In simple terms, a CMUT cell is a capacitor made from a membrane of some material suspended over a vacuum-sealed cavity. Electrostatic force causes the membrane to oscillate over the cavity in response to an applied variable electric field to generate ultrasound, and, inversely, the membrane oscillates in response to an applied pressure signal to generate an alternating current due to the time-varying change in capacitance. Two electrodes are required for electrostatic actuation; one electrode is on top of the membrane (or is the membrane itself), and one electrode is at the bottom of the cavity. Typically, to prevent shorting between electrodes, an insulating layer is located in between the bottom electrode and the cavity. A visualization of this basic structure is shown in figure 4.1. Each of these components can be made from many different kinds of materials, and further discussion on the materials and the methods of fabricating CMUT structures is given in section 4.3. The membrane itself can have many different shapes. CMUTs with rectangular [39], circular [40] and hexagonal membranes

[41] have been fabricated. CMUT cells are then combined into different-sized elements and arrays by defining the electrode shapes, and these used to develop transducers.

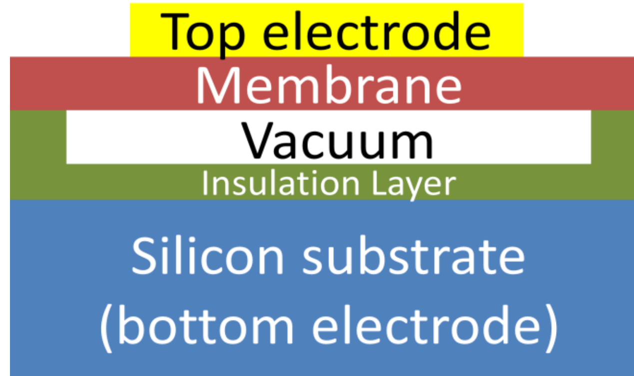


Figure 4.1: Diagram of a typical CMUT cell

4.1.2 CMUT Operation

The basic principle behind CMUT operation is electrostatic force. Given two parallel plates with plate area (A), plate separation (d), the electrostatic force between these two plates due to an applied voltage (V) between them is given by the equation:

$$F_e = \frac{d}{dx} W_e = \frac{d}{dx}\left(\frac{1}{2}CV^2\right) = \frac{\varepsilon A V^2}{2d^2} \tag{4.1}$$

where ε represents the permittivity of the material in between the two plates. In CMUTs, if the top electrode is the membrane and the bottom electrode is not separated by an insulating layer from the cavity, then this value is simply ε_0, the permittivity of free space, because the cavity is a vacuum. If, on the other hand, the electrode is on top of the membrane and there is an insulating layer, the total series capacitance will have to be used in the equation. Alternatively, to account for adding extra layers with thickness d_l and relative permittivity ε_{RL}, 'effective' plate separation d_{eff} can be used in equation (4.1): where

$$d_{eff} = \frac{d_{l1}}{\varepsilon_{RL1}} + \frac{d_{l2}}{\varepsilon_{RL2}} + \cdots d \tag{4.2}$$

This can be used to account for having an insulating layer and an electrode on top of the membrane. For future equations, to maintain simplicity, equation (4.1) will be used with the permittivity of free space. When a voltage is applied to a CMUT cell, the electrostatic force causes the mobile membrane to deflect towards the bottom electrode, which is stationary. Immediately, a problem becomes evident: equation (4.1) is given for a parallel plate capacitor, but as the membrane deflects, because the membrane is clamped at the sides, it no longer resembles a parallel plate. This means that the force on the membrane is not governed exactly by the given equation as it deflects. However, the exact and best way to model CMUT behaviour is a complicated topic that is the subject of significant research in the

CMUT field. Many different finite-element based numerical models have been proposed to accurately predict CMUT behaviour, but for the purposes of this thesis, analytical approximations are used to provide a sufficient background on the operating principles from which it is possible to determine the most important parameters of a CMUT.

As a CMUT membrane is deflected by an electrostatic force, it acts like a mass attached to a spring, generating a mechanical restoring force in the opposite direction which increases as the membrane deflects. Upon the application of a DC voltage to a CMUT, the displacement of the membrane will be at a point where the electrostatic force due to the applied voltage is equal to the restoring spring force. For an out of plane displacement of the membrane (x), the electrostatic force is given by:

$$F_e = \frac{\varepsilon_0 A V^2}{2(d-x)^2} \tag{4.3}$$

where in this case, d represents the initial gap height. The restoring force on the membrane is given by:

$$F_s = k_s x \tag{4.4}$$

where k represents the linear spring constant of the membrane. At equilibrium, $F_e = F_s$, hence the displacement of the membrane can be given as a function of the applied voltage by the formula:

$$V = \sqrt{\frac{2k_s x}{A\varepsilon_0}}(d-x) \tag{4.5}$$

This can be rearranged to form a quadratic equation in x, and solved to find the equilibrium position. Because electrostatic force increases quadratically with displacement, but the restoring force increases only linearly, for some applied voltage, the electrostatic force will overwhelm the restoring force and the membrane will collapse. This voltage is known as the collapse or 'snapdown' voltage. Past this voltage, equation (4.5) no longer has a solution. At the collapse voltage, the derivative of equation (4.5) with respect to x is equal to zero. This occurs at a displacement value of: $x = \frac{d}{3}$. Substituting this value into (4.5) yields a collapse voltage of:

$$V_c = \sqrt{\frac{8k_s d^3}{27\varepsilon_0 A}} \tag{4.6}$$

If the membrane is collapsed, reducing the voltage will eventually cause it to snap back from its collapsed state. The voltage at which this occurs is called the snap-back or pull-out voltage. This voltage is actually smaller than the collapse voltage, because when the membrane is collapsed, the distance between the two electrodes is much smaller, resulting in a much higher electrostatic force. For the

restoring spring force to surpass the electrostatic force, a smaller voltage would be required to compensate. In the case when the CMUT is fabricated such that the membrane and insulating layer are made with the same material (a common design) with respective thicknesses t_m and t_I, the snap-back voltage is given by:

$$V_{SB} = \sqrt{\frac{2k_s d(t_m+t_i)^2}{\varepsilon A}} \tag{4.7}$$

A series of capacitance vs. voltage curves of one of our fabricated CMUTs is shown in figure 4.2. From this figure we can see the hysteresis that occurs due to the differing collapse (shown in blue) and snapback (shown in red) voltage. These two voltages are extremely important parameters to know because they define two different CMUT operating modes which exhibit completely different behaviour, which will be briefly explained in section 4.1.4.

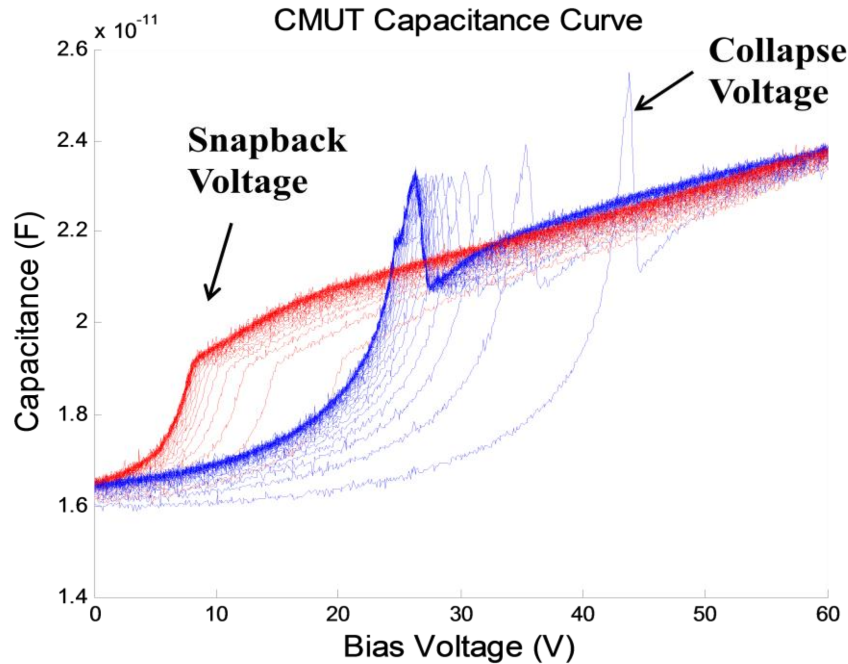


Figure 4.2: Series of capacitance-voltage curves of one of our CMUTs

4.1.3 CMUT Parameters and Equivalent Electrical Model

In addition to the collapse voltage, there are many other parameters of a CMUT which can affect its performance. Although defining the performance of a CMUT depends on properties that can be difficult to extract analytically, a simplified model of operation can be estimated by using the small-signal equivalent circuit model initially proposed by Mason, and shown in figure 4.3. This circuit can be combined with the circuits from interfacing electronics to easily determine transducer frequency response. In this figure, Z_s is the impedance of the source, Z_{med} is the impedance of the medium through which the ultrasound will propagate, and Z_{mem} is the impedance of the transducer membrane. The transformer ratio in the circuit is given by [42]:

$$n = V_{dc} \frac{\varepsilon_0 \varepsilon^2 A}{(\varepsilon_0 t_m + \varepsilon d)^2} \tag{4.8}$$

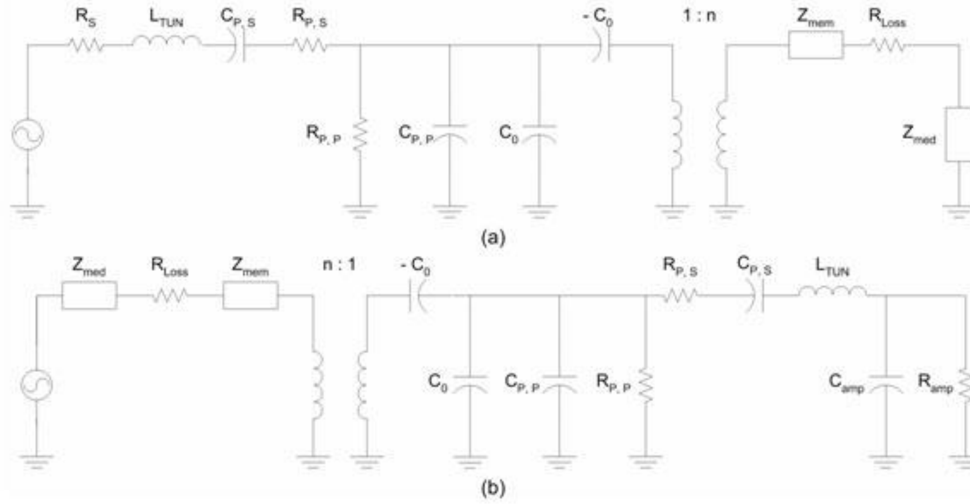


Figure 4.3: a) Mason equivalent model of a CMUT when transmitting ultrasound. b) Mason equivalent model of a CMUT when receiving ultrasound. $R_{S:}$ Source impedance; L_{TUN}: Tuning inductance; C_P, $R_{P,}$: Parasitic capacitances and resistances; C_0: CMUT capacitance; n: Electromechanical conversion factor; Z_{mem}: CMUT Membrane Impedance; Z_{med}: Impedance of transmission medium; R_{loss}: Mechanical loss; C_{amp}: Amplifier input capacitance; R_{amp}: Amplifier input resistance[5]

Calculating the exact impedance of the membrane cannot be done using a simple formula; it depends on the frequency of the applied signal and the membrane geometry, in addition to the properties of the membrane material. The exact impedance, along with other properties of the membrane, can be found by solving the fourth-order differential equation of motion on the membrane which is given by [43]

$$D \left(\frac{\partial^4 w}{\partial x^4} + 2 \frac{\partial^4 w}{\partial^2 x \partial^2 y} + \frac{\partial^4 w}{\partial y^4} \right) = P(x, y) \tag{4.9}$$

where *D* is the flexural rigidity of the membrane given by:

$$D = \frac{E t_m^3}{12(1 - \sigma^2)} \tag{4.10}$$

and *P(x,y)* is the distributed load on the membrane. Regardless of how it is calculated, the important concept to take away is that the impedance of a CMUT membrane is very small – comparable to the impedance of air. [42]. This fact

[5] http://www-kyg.stanford.edu/khuriyakub/opencms/en/research/modeling/Circuit_Model/index.html

causes the CMUT membrane to resonate in air at a specific *resonant frequency* based on the properties of the membrane. When the transmission medium is a liquid such as water or oil, the impedance of the membrane can be ignored, since it is much smaller than that of the medium. Therefore, the transducer has no resonant frequency and displays a wide bandwidth. The implications of this will be discussed in a subsequent section. For a circular membrane with radius a, the resonant frequency is given by the formula [44]:

$$\omega_0 = \sqrt{\frac{k_s}{m}} = \frac{10.22}{a^2\sqrt{\rho t_m/D}} \tag{4.11}$$

where k_s is given by the formulas:

$$k_s = \frac{192\pi D}{a^2} \tag{4.12}$$

A damped membrane has a natural resonance frequency defined by:

$$\omega_d = \omega_0\sqrt{1 - \frac{R_b^2}{4k_s m}} \tag{4.13}$$

where R_b is a damping constant. By substituting equation (4.10) into (4.12), we see that the resonant frequency of a circular CMUT is inversely proportional to its radius and directly proportional to its thickness. Equation (4.13) shows that when a CMUT is damped (such as when it is immersed in a liquid), its natural frequency will *decrease* compared to the resonance frequency in air. Even though the analytical expressions for membrane shapes that are not circular are more difficult to find, the conclusions from the above formulas are still valid. It is evident from these equations that there is a large amount of flexibility and control in designing CMUTs to give specific parameters. The resonant frequency of membranes is defined by both membrane size and thickness, and since the thickness of a single layer cannot readily be manipulated during deposition, it is possible to manufacture membranes of different frequencies at the same time simply by varying the size of the membrane. The collapse voltage and sensitivity of the CMUTs can also be controlled independently from the resonant frequency by varying the gap height.

Another phenomenon that can be derived from the circuit model is called the 'spring-softening effect'. Based on the derivations done in [44], it can be seen that as applied voltage increases, the effective spring constant of the system approaches zero. This has the effect of decreasing the natural resonant frequency. In addition, as the applied bias voltage is increased, the electromechanical coupling coefficient also increases. As discussed in section 2.3 for piezoelectric transducers, it is desirable for this coefficient to be as high as possible, as it represents the efficiency with which the transducer converts electrical energy into mechanical energy and vice versa. The same logic holds for CMUTs, where the coefficient is given by [45]

$$k_T^2 = 1 - C(x)/(\frac{d}{dV}\left(VC(x)\right) \quad (4.14)$$

where *C(x)* is the capacitance of the membrane at a displacement of *x*. Therefore, a biased CMUT will operate at a lower natural frequency than an unbiased one, but with a higher efficiency. An example of the electromechanical efficiency vs. voltage curve for one of our CMUTs is shown in figure 4.4, from which it can be confirmed that the most efficient point of operation of a CMUT when it is biased at nearly the collapse voltage.

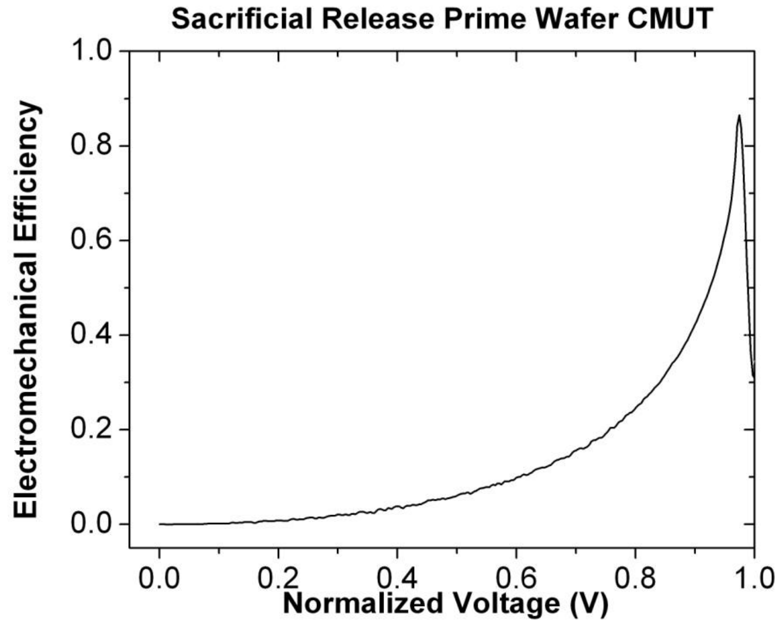


Figure 4.4: Electromechanical efficiency vs. voltage curve for one of our CMUTs

It is important to note that the formulas used above do not include a term for residual stress on the membrane. During fabrication, the material deposited to form the membrane can have some inherent stress in its plane. This stress value changes the value of the spring constant, and in some cases, can compromise the membrane structure. For plates, a tensile (pulling) stress increases the stiffness (spring constant) of the plate [43], and would thus result in a higher than expected resonant frequency whereas a compressive stress reduces the stiffness, and would thus result in a lower than expected resonant frequency. Too much tensile stress, however, can cause tearing in the membrane, and too much compressive stress, on the other hand, can cause buckling of the membrane. In this latter case, the equilibrium position of the membrane is either collapsed or inflated upwards, which would make it impossible to obtain effective membrane actuation. It is therefore important to try and reduce stress as much as possible during CMUT fabrication.

4.1.4 CMUT Operating Modes

As previously mentioned, the collapse voltage and snapback voltage define different operating modes for CMUTs. A CMUT cell can actually operate in three distinct modes, each with its own particular characteristics and applications. The first is known as pre-collapse mode and is the conventional mode wherein the membrane and the substrate do not make contact. In this mode, the electromechanical coupling coefficient increases as the voltage nears the collapse voltage, and it is therefore advantageous to operate as close to the collapse voltage as possible, since the membrane will vibrate with greater displacement for an applied AC voltage signal. A second mode of operation is called collapse mode operation. In this mode, the membrane is intentionally collapsed and then actuated. Because the center of the membrane is contacting the bottom of the gap, only the un-collapsed portions of the membrane vibrate, effectively decreasing the membrane area and thus significantly increasing the frequency of ultrasound generated. In tests performed with our TOBE CMUTs, collapsing the membrane caused the resonance frequency to increase from 8.5MHz to ~22MHz. There has been significant research done to understand and analyze collapse-mode operation; other groups have found that in this mode, linearity improves and the electromechanical coupling coefficient increases. One of the disadvantages of operating in this mode is the prevalence of charging, which has been investigated in [46]. As charge builds up in the membrane and insulating layer, it can degrade the performance of the membrane, or cause it to remain collapsed due to the establishment of a voltage potential from the stored charge. Figure 4.2 demonstrates the effects of charging on the collapse voltage, causing it to shift after repeated membrane collapses. The shift of the C-V curve also results in a shift of the electromechanical efficiency curve, potentially decreasing the efficiency of CMUT operation when biased. However, we also found that charging can potentially increase the operating efficiency at low biases [47]. For the tests performed in the course of my research, pre-collapse mode was used almost exclusively, as we did not find the same performance improvements when operating in collapse mode. A third, less studied mode is known as collapse-snapback mode. In this mode, the applied AC voltage signal has a high enough amplitude to snap the membrane down, and then allow the membrane to snap back up. While this mode has high membrane displacements, resulting in high output pressures, it is also the least linear, and difficult to predict. Visualizations of pre-collapse and collapse mode operation are shown in figure 4.5.

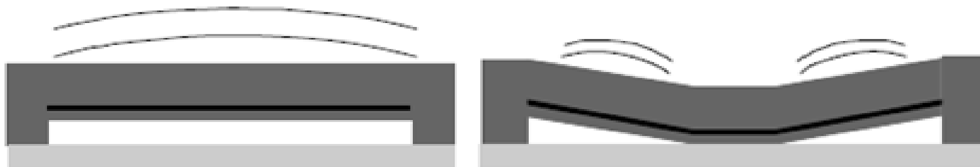


Figure 4.5: Operation of a CMUT in pre-collapse mode on the left, and collapse mode on the right [48] © 2009 IEEE

4.2 Comparisons with piezoelectric transducers

Because piezoelectric technology was developed much earlier than CMUTs, it had already established itself as the industry-leading technology even when research into CMUTs was still in its infancy. It is therefore necessary to make a direct comparison between CMUTs and piezoelectric technology before drawing conclusions on the potential for CMUTs to become a dominant technology.

4.2.1 Advantages of CMUT Technology

Although CMUTs did not immediately invade the well-established piezoelectric transducer market upon their inception, significant advantages over the latter were immediately obvious, both in terms of individual element operation and integration into arrays. The clearest advantage of CMUTs over piezoelectric transducers was their operation in air. In fact, the initially proposed applications for CMUTs were limited to air-coupled applications. As mentioned in section 4.1.3, the impedance of the membrane material is very small: on the same order as that of air. By comparison, the impedance of one variation of PZT is 33.7MRayl [3]. This is much higher than even the impedance of tissue (~1.5MRayl). From the concepts discussed in section 2.1.3, it is clear that this large impedance mismatch between piezoelectric transducers and the transmission medium results in inefficient ultrasound transmission into the medium; much of the ultrasound is lost due to boundary reflection. Matching layers are thus needed for piezoelectric transducers to image effectively. CMUT transducers do not require these matching layers because of their much more closely matched impedance, making these transducers more suitable for such applications as NDT, microphones, or other air-coupled applications.

Despite having initially been designed for air-coupled use, CMUTs also displayed advantages when used in immersion. Because the acoustic impedances of water and other liquids are much higher than those of CMUT membranes, the impedance of the membrane can effectively be ignored in immersion operation. This results in an elimination of resonance frequency effects for these transducers, and consequently a drastically higher bandwidth. Results have shown that in immersion testing, CMUTs can have as good a dynamic range as a piezoelectric transducer with a flat frequency response over a wide range [42]. Specifically, CMUTs that have more than 100% fractional bandwidth have been demonstrated [49]. This broadband, non-resonant response allows the transducer to emit and detect many more frequencies contained in a short ultrasonic pulse, resulting in better depth resolution [2]. Thus with their wider bandwidth, CMUTs displayed the potential for superior imaging quality.

Another advantage that CMUTs have over piezoelectric transducers is that the micromachining methods used in the fabrication of CMUTs are not only comparatively easy and allow for high-yield batch processing, but they also allow for a much greater flexibility in the integration of CMUTs into different kinds of arrays. As discussed in detail in section 3.3, for fully-wired 2-D arrays, the maximum element pitch is restricted to less than half the acoustic wavelength at the operating frequency in order to prevent the appearance of grating lobes. The

‘dice and fill’ fabrication process of piezoelectric arrays is difficult to apply to densely packed arrays. However, with CMUTs, the use of simple photolithography makes it comparatively easy to define complex 1D or 2-D arrays. These arrays can be easily integrated with their associated transmit/receive electronics using a number of different methods which will be discussed in section 4.4. The ability to have integrated electronics, along with the thin profile of CMUT arrays (equal to the thickness of the silicon wafer) allows arrays to fit in tight spaces. This could allow CMUTs to be used in endoscopic applications in combination with other imaging techniques such as photoacoustic imaging [50], which requires extremely sensitive detection, or implantable sensors.

CMUTs also allow for greater flexibility in array design, which can result in more specialized ultrasound transducers. For example, ultrasound has been used for both clinical imaging systems and therapeutic applications such as high intensity focused ultrasound for thermal killing of cancers [1]. Therapeutic applications typically require more intense, low-frequency ultrasound, whereas imaging requires higher frequency and higher bandwidth ultrasound to obtain better resolution. Because the frequency of piezoelectric transducers is defined by the thickness of the material layers, it is not readily possible to satisfy both these requirements in the same fabrication process. However, different frequency CMUTs can easily be integrated on the same device by interlacing CMUTs with varying membrane sizes together. An example of such an interlaced array designed and fabricated by our lab group is shown in figure 4.6.

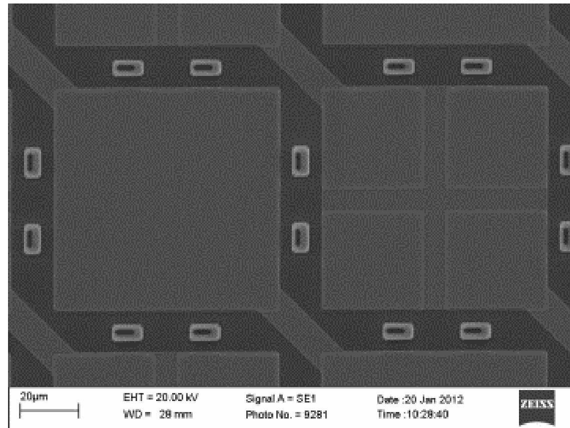


Figure 4.6: SEM image of a dual-frequency CMUT array

Overall, CMUT technology has come far since the initial days when it could not compete with piezoelectric transducers. Currently, 128-element 1D CMUT arrays have already been fabricated and used with commercial scanners to obtain clinical images with improved resolution over piezoelectric arrays [23]. In addition, 32x32 element 2-D arrays have also been fabricated and are being used in imaging tests [23].

4.2.2. Disadvantages of CMUT Technology

Despite all the advantages mentioned in the previous section, CMUT technology did not become prevalent in imaging equipment because of the initial problems that they displayed. One primary disadvantage of CMUT technology is that, in comparison with piezoelectric transducers, CMUTs have very low device capacitance, making them more vulnerable to parasitic capacitance. In 2-D arrays especially, the capacitance is so small that connecting any cable to the CMUT is detrimental to the transducer operation [51]. However, the aforementioned ability to integrate the required electronics on the same wafer as the CMUT transducers has since mitigated this problem.

Some of the more serious problems that are mainly responsible for the slow adoption of CMUT technology are more difficult to resolve. These problems involve long term reliability issues with the CMUTs and present a significant commercial barrier. The most significant of these problems involve charging. As discussed in section 4.1.4, charges that become trapped in the membrane and insulating layer can cause unpredictable changes in the operation of the CMUT [52] and can even prevent the membrane from snapping back after collapse, making it necessary to keep adjusting the DC bias to compensate [2]. Replacing the isolation layer with several individual isolation posts was proposed as one way to not only significantly decrease the effects of charging by reducing the volume of dielectric, but also eliminate hysteresis in the collapse-mode operation of the membrane, with no loss in performance [52]. Operating in pre-collapse mode can also mitigate charging effects. The nonlinearity of CMUT response to applied AC voltage signals can also make CMUT operation more difficult to model, and can become problematic when imaging the nonlinear response of biological tissues [2]. Purposely distorting the drive signal to compensate for this nonlinearity can alleviate this issue.

The common theme with many of the disadvantages is that an increased amount of research into CMUT technology is resulting in solutions to these problems. It is highly likely that it will not be long before the theoretical advantages of CMUTs will materialize into their actual adoption in industry and replacement of piezoelectric technology. The use of CMUT technology in niches where it already holds significant advantages over piezoelectrics, and where the latter technology is most critically limited, such as complex 2-D and multi-frequency arrays, could occur even sooner.

4.3 CMUT Fabrication

Throughout the decades of CMUT research, there have been many proposed methods of fabricating these devices. It would be impossible to cover every single process variation, but this section will first serve to give a background on the general types of fabrication processes for CMUTs. The two main approaches to CMUT fabrication: sacrificial release, and wafer bonding, will be presented and compared. The work done and observations with my own fabrication work will also be included.

4.3.1 Sacrificial Release Process

The sacrificial release process was used to fabricate the very first CMUTs and was used for many years successfully as the standard fabrication method [51]. The basic process flow behind the sacrificial release method is shown in figure 4.7. The process begins with a silicon wafer which is highly doped. The highly-doped wafer is made to be conductive so as to act as the bottom electrode in the CMUT. The next step is to create the insulation layer as in figure 4.7 a). While the initial CMUTs described in [53] and [54] did not include this layer, future iterations of this method such as those in [51] included this layer to remove the possibility of shorting. The insulation layer can also act as an etch-stop if the sacrificial layer is made out of a material that can be etched with high selectivity. Because the insulation layer is the first layer over the silicon wafer, it can either be grown by oxidation (if SiO_2 is the desired insulation material), or deposited by chemical vapour deposition (CVD). The properties of this layer are important in defining both device performance and the choice of materials for further fabrication steps. As previously mentioned, dielectric charging in the insulation layer can pose a significant problem in CMUT operation, and making the insulation layer as smooth as possible minimizes this effect. Also, poor quality layers, especially thin Si_3N_4 layers [51] can often contain pinholes, especially if the insulation layer is not sufficiently thick. Pinholes can cause problems because during isotropic etching, they can allow an etchant to seep through the insulation layer and reach the substrate, potentially etching it. They can also lower the breakdown voltage of the insulating layer. Typically, a layer over 100nm thick is necessary to avoid them. For this reason, better quality deposition methods such as thermal oxidation or low-pressure CVD (LPCVD) are preferable for this step. However, low temperature deposition methods such as plasma-enhanced CVD (PECVD) can allow for a CMOS compatible process, which allows the CMUTs to be fabricated on top of pre-fabricated electronics. Requirements for CMOS-compatible CMUT fabrication processes are further discussed in section 4.4.

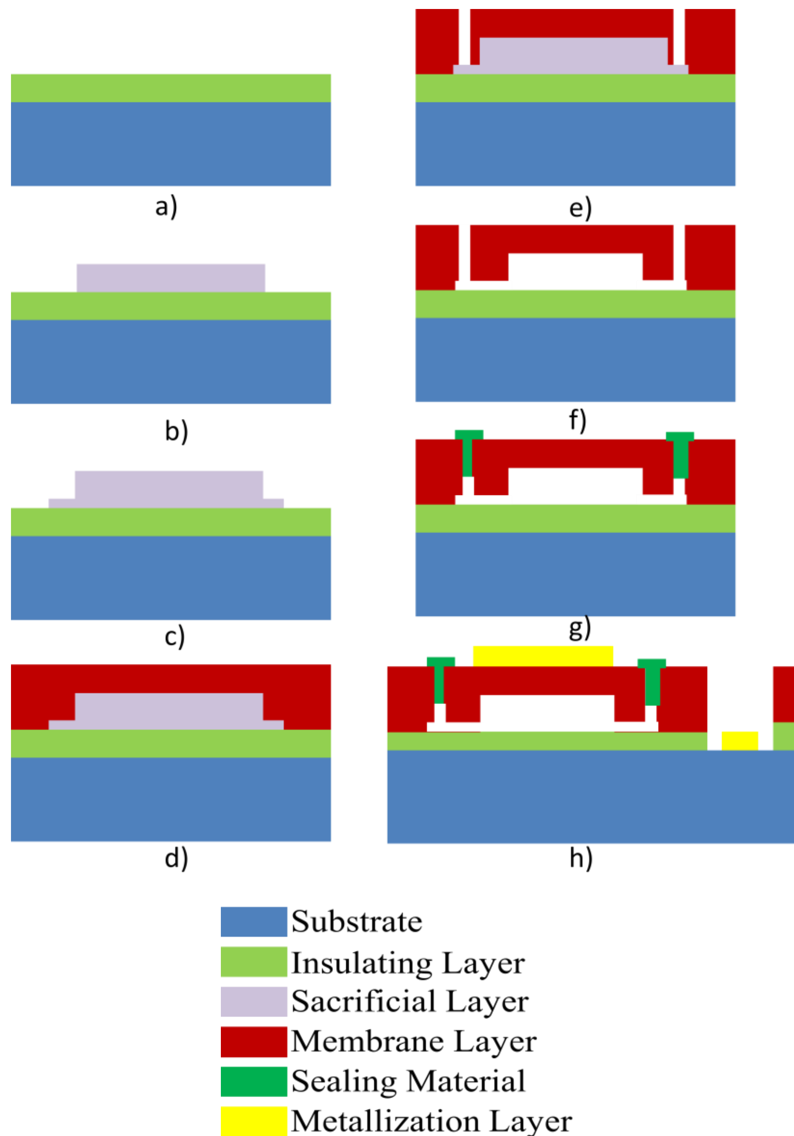


Figure 4.7: Process flow for a sacrificial release CMUT fabrication process: a) Deposition of insulating layer; b) Deposition and patterning of initial sacrificial layer; c) Deposition and patterning of second sacrificial layer to define cavity; d) Deposition of membrane layer; e) Etching of etch holes f) Sacrificial release step to release membrane; g) Deposition and patterning of sealing plugs; h) Deposition and patterning of metallization layer after etching access holes

The next step is the deposition of the sacrificial layer. There are two steps to the deposition and definition of this layer: The first deposition and patterning step (figure 4.7 b)) defines the membrane area, and the second step (Figure 4.7 c)) defines the region known as the etch channel, through which the sacrificial layer etchant will reach the cavity. CMUT membrane sizes are large enough that conventional photolithography can be used during patterning to obtain suitable resolution. Because vertical sidewalls are desired for the membrane and etch channel walls, an anisotropic dry etch procedure such as reactive ion etching

(RIE) is preferable for these patterning steps. The quality of the sacrificial layer is not as important as that of the insulation layer, and it can just as effectively be deposited at low temperatures, since it plays no part in the final CMUT. The most important requirement for this layer is that it can be etched with good selectivity against the membrane and the insulation layer. Figure 4.8 demonstrates what can occur if the selectivity between the membrane and the sacrificial layer is poor; this effect can cause device operation to deviate significantly from expectations. Depending on the size of the membranes, the sacrificial layer may also have specific stress requirements. This will be discussed subsequently.

Figure 4.8: Diagram depicting the consequence of poor etch selectivity during sacrificial layer etch

The next step is the deposition of the membrane (Figure 4.7 d)). Because the membrane is the moving structural component of the CMUT, the mechanical properties must be well defined and characterized, and hence it is desirable to have a high quality layer with little variation in thickness throughout the whole wafer, again favouring high temperature deposition processes. One important property of the membrane that must be controlled is stress. As shown in [55], and previously discussed, high compressive stresses will cause a membrane to buckle after release and thus render the CMUT unusable. Thus it is desirable to use a deposition method and a membrane material that will result in a film with low tensile stress. For example, single-frequency PECVD systems that use tetraethyl orthosilicate (TEOS) to deposit silicon based layers, such as the one at the University of Alberta NanoFab, cannot directly deposit tensile-stress oxide membranes. It has been found by other groups that a PECVD using silane based chemistries can deposit achieve this [56]. Dual frequency PECVD systems can also allow for better control of thin film stress levels during deposition [57].

After membrane layer deposition, the etch holes through which the sacrificial layer etchant will penetrate the membrane layer are patterned and etched using a dry etch step (figure 4.7 e)). Because these holes have a high aspect ratio, it is important that an anisotropic etch like RIE is used. Following this, the wafer is immersed in the sacrificial layer etchant (which must be an anisotropic etchant so as to be able to etch underneath the membrane) and the membrane is released (figure 4.7 f)). As mentioned previously, it is imperative that this etchant has very high selectivity between the sacrificial layer and membrane layer. Differences and comparisons between different etchants will be discussed in the next section. To seal the CMUT, the etch holes must then be plugged (figure 4.7 g)). This is done by depositing an additional thick layer on top

of the membrane layer. This layer can either be combined with the previously deposited membrane layer to form the total thickness of the membrane (if it is the same material as the membrane), in which case no additional etching step is necessary, or it can be patterned so as to cover only the etch holes as in the process flow diagram, in which case additional lithography and etching is required. If such is the case, then both dry and wet etching can be used; however, dry etching can roughen the surface of the membrane and lead to increased charging or physical membrane etch, whereas wet etching can undercut the photoresist and completely etch the sealing plugs away. During fabrication of our CMUTs, it was found that a combination of dry etching and wet etching worked best, with dry etching being used to etch away most of the sealing layer, then wet etching being used to etch the leftover section. In such a way, roughening of the membrane was avoided, and undercut of the photoresist over the sealing plug was minimal. The sealing is an important step because it prevents the penetration of fluid such as air or water in the membrane cavity, which would impede CMUT operation. This means that the deposition of the sealing material must occur in a good vacuum. A material and a deposition method with a high sticking coefficient will ensure better sealing, since it will plug the holes without coating the inside of the cavity. For such a reason, LPCVD would not be a suitable sealing method, whereas a process such as evaporation, which has a highly vertical deposition profile, would be ideal.

The final steps define the metal contacts to the top and bottom electrodes. First, the holes which allow contact to the bottom electrode (the doped wafer) are patterned and etched using either wet or dry etching (if the membrane and insulation layer are made out of the same material, this can be done in one step). Metal is then deposited on the wafer, patterned, and etched to define the metal contacts (figure 4.7 h)). In the literature references provided up to this point, aluminum or gold has been used as the contact material. Because the metal contacts do not need to be thick (<200nm) and have a very low aspect ratio, isotropic etching can, and often is used to etch the metal layer, since many metals do not have suitable dry etchants. After removing the photoresist mask, the sacrificial release method is complete, and the result is the finished device.

4.3.2 Sacrificial Release Method Process Variations

While the process flow described above is the basis of all sacrificial release methods, many different variations have been explored. The biggest source of variation is the materials used in the three initial layers – the insulation layer, the sacrificial layer, and the membrane layer. Changing the materials in these layers also changes the required etchants as well, since the different materials have varying selectivity to certain etchants. Generally, membranes made out of silicon-based material such as SiO_2 or silicon nitride are favoured due to the mechanical properties of silicon [2]. The initial process in [53] used wet grown oxide as the sacrificial layer and LPCVD nitride as the membrane layer. HF was used to etch the oxide and release the membrane. While this process can successfully release the membrane, it is not ideal. The selectivity of HF between SiO_2 and nitride is only around 100:1 [51] which is not enough to ensure membrane thickness

uniformity. The most popular design seen in many sources of literature such as [51] and [58], uses a nitride insulation layer, silicon (poly or amorphous) sacrificial layer, and an LPCVD nitride membrane. The etchant used for our initially fabricated CMUTs was KOH, which has a selectivity of over 400000:1 [51] between nitride and silicon. Other suitable wet etchants would be HNA (a combination of HF/Nitric Acid) or EDP (an aqueous solution of ethylene diamine and pyrocatechol) due to their low SiN etch rates [59]. All three of these etchants do slowly etch oxide, so a nitride insulation layer would be more ideal. J. Knight in [58] investigates using metal as the sacrificial layer as well. One advantage to using metals is that they can generally be etched selectively to silicon-based materials. However, using certain metals may limit the temperature of the processes used to deposit the membrane layer. Knight concludes that aluminum is a poor choice due to oxidation during the nitride coating process, and that chromium works much better since it does not oxidize in the PECVD chamber.

One factor that is present with all the variations presented above is that they use wet etchants to etch the sacrificial layer. Wet etching presents the problem that during the drying phase, surface tension can cause the etchant to adhere to the membrane and pull it down as it evaporates, causing the membrane to become stuck to the insulation layer (sticktion). These processes thus require a supercritical drying or freeze-drying step to prevent such an occurrence. Isotropic dry etching can resolve these problems. Noble et al. proposed using a polyimide sacrificial layer with a nitride membrane in [56] that would require a dry oxygen plasma etch to release. This architecture would have very good selectivity since oxygen plasma does not etch silicon. One dry etchant that has seen limited use in previous literature is XeF_2, which has a very fast silicon etch rate. This etchant is described in more detail in the next section.

Another factor which must be taken into account is the stress of the sacrificial layer. Layers with high compressive stress, such as LPCVD polysilicon, when released by etching, can cause larger membranes to break [51], limiting the maximum size of the membranes. This is a problem with the popular nitride-poly-nitride architecture and was one of the motivating factors to move towards a wafer-bonding fabrication process. Overall, there are countless possible combinations of materials that can be chosen for the sacrificial release technique, and they cannot all be listed. No matter what materials are chosen, however, they all share the requirement that the sacrificial layer must have an etchant which is selective against both the membrane and the insulation layer.

4.3.3 Sacrificial Release Fabrication Observations and Results

A number of different sacrificial release based processes were attempted during the course of this degree to establish the best possible method to fabricate CMUTs. Initially, CMUTs were fabricated using the process shown in figure 4.7. The insulating layer was stoichiometric silicon nitride (Si_3N_4) and the sacrificial layer was polysilicon. The membrane was a composite structure composed of a thick layer of low-stress, silicon-rich nitride sandwiched between two thin layers of stoichiometric nitride. The low stress material proved more suitable for the membrane because of its relatively low tensile-stress, but the thin layers of

stoichiometric nitride were added to provide better etch selectivity against the polysilicon sacrificial layer. KOH was used as the sacrificial release etchant, SiO_2 was used for the sealing plugs, and aluminum was used for the metallization layer. Upon the completion of the process, it was observed that many of the larger membranes, with dimensions of (116 μm)² were either snapped down due to sticktion or broken due to the large stress of the sacrificial layer. The TOBE CMUTs, which had dimensions of (42 μm)², did not suffer from these effects.

To resolve these problems, numerous variations of a dry-etch process were also attempted using XeF_2 as the sacrificial release etchant. Both CMOS-compatible and high-temperature processes were attempted. Because XeF_2 is very selective against oxide, an oxide membrane would be the ideal choice over a nitride membrane, since the latter has a non-negligible etch rate in XeF2 [60]. However, obtaining a low-tensile stress oxide membrane proved challenging. PECVD oxide was used as a first attempt, and annealing was used to try and remove the compressive stress, which was not easily controllable during the deposition and varied from -200MPa to -50MPa. Annealing the high-compressive-stress film was able to reduce the stress in some cases, but the final stress was not easily predictable and generally did not end up being tensile. At some temperatures, annealing even caused the compressive stress to increase. It was found that an inductively-coupled plasma CVD (ICPCVD) located at the Indian Institute of Technology in Bombay could deposit low tensile-stress oxide, but upon receiving the wafers and performing sacrificial release etching at the University of Alberta, buckling was still observed. This was likely due to an increase in the compressive stress of the oxide film over time, a phenomenon which is supported by published literature [61].

In addition to the stress and membrane buckling issues, problems were also observed with the sacrificial release process. For the CMOS compatible process, low-temperature amorphous silicon (a-Si) was used as the sacrificial release layer. Both ICPCVD-deposited a-Si and sputtered a-Si were used. Although the etch rate of oxide is negligible in XeF_2, significant etching of the membrane and substrate (due to the etch-through of the thermal oxide insulating layer) was observed during the processes. After much study, it was determined that the oxide etch was due to dangling hydrogen bonds on the a-Si resulting in a-Si:H, which can react with the fluorine in XeF_2 to produce HF gas. This gas in turn can etch oxide. A similar effect can be observed when there is excess water vapour remaining in the etching chamber or on the wafer, since the vapour will also react to form HF gas [62]. Similar observations were not made when LPCVD-deposited polycrystalline silicon was used as the sacrificial layer, however, the use of this deposition method negated CMOS compatibility, and did not resolve the compressive stress issues with the oxide membrane. Because no available oxide deposition techniques could provide tensile stress oxide layers, LPCVD nitride was instead used with a thin protective PECVD oxide layer underneath. This method proved to be successful, and no breakage or sticktion of large membranes was observed after sacrificial release. Testing of these membranes to compare their performance to the previous successfully used method will be a point of future experimentation.

4.3.4 Wafer-bonding Process

While silicon bonding is an old technology, it was only expanded more recently for use in CMUT processing, having first been developed in the early 2000's [63]. Because wafer bonding uses two wafers – one to form the membrane and one to define the cavity – this method intuitively offers more flexibility in CMUT architecture design than the sacrificial release method since the design parameters of the membrane and the cavity can be separately optimized. Consequently, the variations in the wafer bonding processes extend beyond using different materials for the thin film layers and generally have structural differences. Several of these variations will be presented. To give an overview of the process, the process flow for the fabrication of the simplest type of wafer-bonded CMUT will be detailed (shown in figure 4.9) and then the differences present in some of the other variations will be mentioned.

The basic start for any wafer-bonding process is two wafers, one of which is an SOI wafer. In the simple process, the second wafer is a doped prime quality silicon wafer. Oxide is first grown on this wafer to act as the cavity walls (figure 4.9 a)). The thickness of this oxide will be equal to the desired thickness of the cavity. The membrane area is then defined using photolithography and the oxide is etched to form the cavity, with the silicon wafer acting as the etch stop (figure 4.9 b)). Because in general, the CMUT size is much greater than the thickness of the oxide layer, the etching can also be done with buffered oxide etch (BOE) without worrying about significant undercut. In addition, BOE prevents roughening of the silicon wafer, which can result in a rough insulating layer and increased charging. After the oxide-etch, another thinner layer of thermal oxide is grown to act as the isolation layer (figure 4.9 c)). Because the isolation layer is a thermally grown oxide, it has much better quality and suffers less from dielectric charging than isolation layers deposited by CVD. In fact, Lin et al. characterized the surface roughness of various thin films using an atomic force microscope in [64] and found that thermal oxide had a thickness variation of only $\pm$0.2nm compared to $\pm$ 1.2-1.5nm for LPCVD nitride. LPCVD polysilicon, which is often used as the sacrificial layer in the sacrificial release method, has a thickness variation of $\pm$23.9nm. Using thermally grown oxide result will thus also result in much better gap height uniformity throughout the wafer. The SOI wafer is then fusion bonded to the prime wafer with the silicon device layer on the bottom and the bulk substrate on top (figure 4.9 d)). One requirement for silicon fusion bonding is that the wafer must have a near-atomically flat surface to obtain successful bonding. This typically means that plasma processes (such as RIE) should not be used on the bonding layer, because these processes can significantly roughen the surface. Once properly bonded, the bulk substrate and oxide layers are then removed to leave the silicon device layer as the membrane (figure 4.9 e)). This is done by first grinding the silicon substrate and etching the rest using an etchant such as KOH or TMAH, and then using BOE to remove the buried oxide layer.

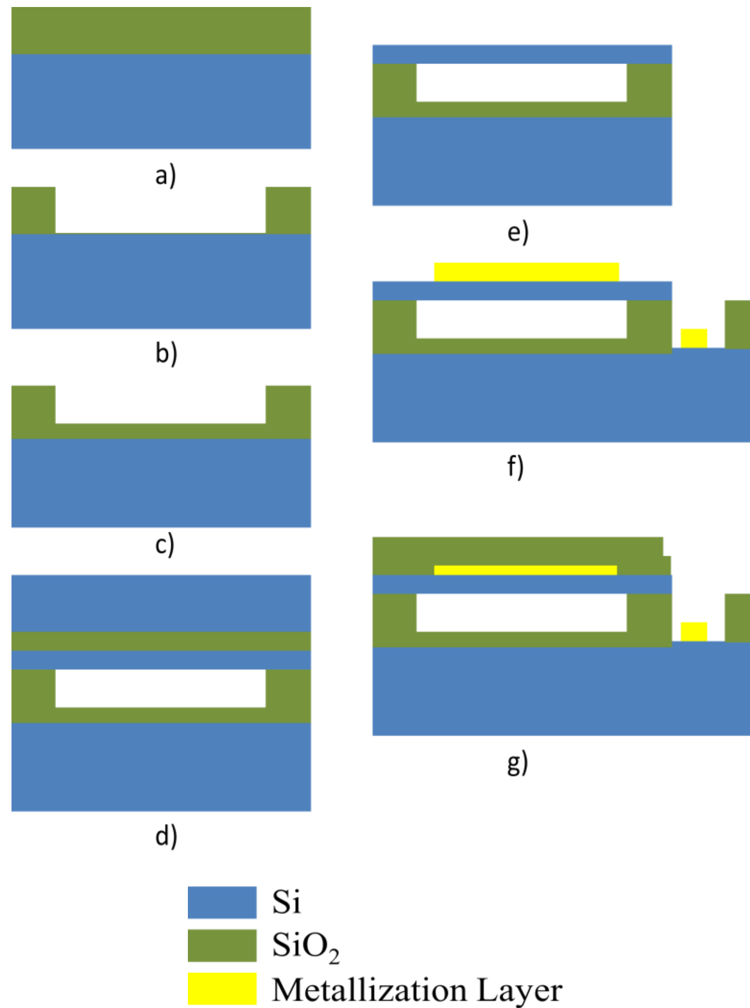


Figure 4.9: Process flow for a simple wafer-bonding fabrication method: a) Thermal oxide is grown on a silicon wafer; b) CMUT cavity is etched in the oxide; c) Thermal oxide is re-grown to act as insulation layer; d) SOI wafer is fusion bonded to Si wafer; e) Handle and buried oxide are removed; f) Metallization layer is deposited and patterned; g) Additional passivating layer is deposited if necessary

The penultimate step is the patterning of the top membrane using photolithography and the etching through the membrane and thermal oxide layer to create holes such that the bottom electrode can be accessed. This step can also be used to create isolation trenches between devices. Because at this point, bonding has already been successful, using RIE is acceptable. Finally, the metal electrodes and contacts are sputtered and patterned similarly to how this was done in the sacrificial release method (figure 4.9 f)). A passivating layer can be deposited to act as insulation, but this then requires access holes to the contact pads to be patterned and etched to complete the device (figure 4.9 g)). For silicon membranes, especially those that are doped, an insulating layer can prevent shorting when wirebonding to the bottom electrode, since the wires would not make direct contact with the membrane.

4.3.5 Wafer-Bonding Method Process Variations

Unlike the sacrificial release method, the majority of variations in the wafer-bonding process are structural variations as opposed to material variations. This is because not only is the membrane material limited by its necessity to bond with silicon, but the single-crystal silicon device layer of the top wafer has some of the most suitable structural properties for use as a membrane, so there is little incentive to replace it. A common variation of the wafer-bonding method is to use an SOI wafer as the bottom wafer, as presented in [65] and [39]. The device layer of the bottom SOI wafer is then used as the bottom electrode. With this structure, trenches can be etched to isolate sections of the device layer, as shown in figure 4.10. This reduces parasitic capacitance by restricting the bottom electrode area entirely to the region underneath the CMUT cavity. In this design, the gap is defined either by RIE of the silicon device wafer or thermal oxidation and etch back of the oxide (which yields better uniformity). Deep RIE can be used to etch the insulation trenches. Other variations, such as the example in [66], involve implementing a piston structure into the membrane. The piston structure is intended to improve CMUT performance by creating a more uniform surface displacement profile, but creating it uses three bonding steps instead of two.

Figure 4.10: SOI wafer with trenches etched into silicon device layer. The silicon device layer is heavily doped to act as a bottom electrode

4.3.6 Process Comparisons

Overall, the wafer-bonding process has proven to display numerous advantages over the sacrificial release method. Firstly, the membrane quality is much better in wafer-bonded devices, owing to the fact that it is made out of a single crystal plate with no residual stress and little thickness variation or roughness across its surface. The mechanical and electrical properties of single-crystal silicon are well known. The improved uniformity in gap height due to the thermally grown oxides and single crystal silicon membrane minimize dielectric charging. Additionally, the wafer-bonding process is also much simpler, and thus results in higher yield. In the original simple wafer-bonding process, there are only four lithography masks needed instead of six, and only two oxidation steps instead of five depositions. As such, the processing time is significantly reduced for the wafer-bonded CMUTs. While the cost of the multiple SOI wafers required is greater than the cost of a single prime wafer, the savings in processing costs can offset this extra cost for the wafers.

4.4 CMOS Integration and Packaging

As previously mentioned, the capacitance of an individual CMUT element is small enough that connecting it to its required transmit/receive electronics by

any cable results in detrimental parasitic capacitance. Thus, connections between CMUTs and CMOS electronics require more novel integration methods. This section will not cover the fabrication process behind the CMOS electronics themselves, but will rather elaborate on the steps needed to integrate CMUTs with these electronics and what makes a CMUT fabrication process 'CMOS compatible'.

One integration method possible is called monolithic integration, where the electronics and the MEMS devices are both fabricated on the same wafer. This method is in turn divided into three separate methods. These are post-processing, co-processing, and pre-processing. In pre-processing, CMUTs are fabricated on the wafer before the electronics. There has been no literature found documenting this type of integration, most likely because fabricating CMOS electronics on a substrate already containing MEMS devices is quite difficult. In co-processing, electronics and MEMS are fabricated on the wafer simultaneously. Eccardt et al. were among the first to integrate CMUTs with electronics and they did so using a co-processing method [67]. The CMUT fabrication steps were completely integrated into a 16 mask BiCMOS process with only minor modifications such as an additional photolithography step and sacrificial layer etch step. The co-processing method has two major limitations, however. First of all, because the electronics and the CMUTs are fabricated side-by-side, they share the same wafer, limiting the amount of area available for the transducers, which restricts the fabrication of large arrays such as 2-D arrays. Secondly, because the specific dimensions of the process are bound within the confines of the BiCMOS process, it does not allow for control over the gap height, the membrane thickness, or the residual stress of the membrane.

The most popular form of monolithic integration is post-processing, where the CMUTs are fabricated after the finished CMOS electronics. One method of post-processing that has shown to have good control of device dimensions is simply passivating and planarizing the surface of the electronics layer, opening contacts to the electronics using dry etching, and then further depositing thin films to define the CMUT structure using the sacrificial release process. The basic structure is shown in figure 4.11. Post-processing introduces temperature limitations to the CMUT fabrication process. To ensure that no damage is done to the metal in the electronics, the temperature of any post-processing steps must be kept below 400ºC. In addition, there are limitations to the etch processes that can be used. Both KOH and HNA are not CMOS compatible. As described in section 4.3.3, XeF_2 is a desirable sacrificial etchant for a CMOS compatible architecture, and, despite unsuccessful attempts in the NanoFab, could be adapted for such a process given the right tools and materials. The architecture presented by Noble et al. in [56] uses a post-CMOS method with a polyimide sacrificial layer, which can be etched in a CMOS-compatible manner using oxygen plasma.

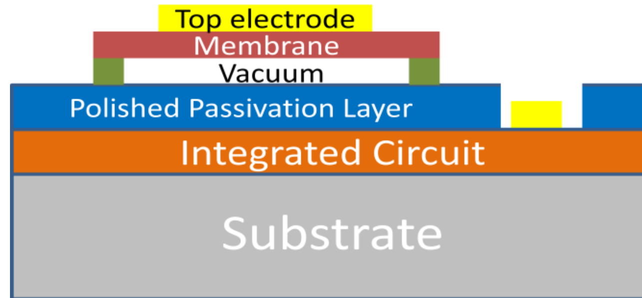


Figure 4.11: Diagram of a post-CMOS integrated CMUT structure

Since silicon fusion bonding is a high-temperature process, there have been much fewer examples of CMOS compatible monolithic integration using wafer-bonding methods. Tsuji et al present a CMOS compatible wafer-bonding method in using a low temperature bonding (400ºC) bonding process with a titanium adhesion layer and using low-temperature oxides as insulation [68]. However, the lack of thermal oxidation results in less control over the gap height and the uniformity of the insulation layer.

Due to the significant advantages offered over the sacrificial release fabrication method, it became necessary to implement an effective way to be able to integrate CMUTs with their electronics while allowing the wafer bonding process to be used. The advantages of monolithic integration in reducing parasitic capacitance are lost when taking into account the sacrificial release process limitations which cause CMUT device performance and yield to suffer. One way to combine the advantages of monolithic integration with the flexibility of isolated CMUT fabrication is to fabricate the CMUTs and CMOS electronics on different wafers and then bond them together using flip-chip bonding, as shown in figure 4.12. For the sacrificial release process, only minor modifications need to be made to enable flip-chip bonding. A thermal oxide is first grown before the entire process to isolate the wafer from a through-wafer via which will connect the CMUT electrodes to the other side of the wafer. The via is then etched using DRIE and the wafer is coated with conductive polysilicon using LPCVD. The conductive poly-Si is etched away everywhere except for within the via, which is then filled with a nonconductive polysilicon filling. The bottom electrode is deposited along with the flip-chip bond pad, and then the sacrificial release process continues as previously described. The CMUT wafer is then flip-chip bonded to the wafer which has the electronics on it. Flip-chip bonding is a well understood process wherein the flip chip pads on the CMUT wafer are aligned with the pads on the CMOS wafer and soldered. Since this process allows independent optimization of the CMUTs and the CMOS electronics, with each process being individually less complicated than a monolithic integration process, it can provide increased yield by comparison. Perhaps its greatest advantage is its compatibility with the wafer-bonding method by simply etching the vias before beginning fabrication on the wafer.

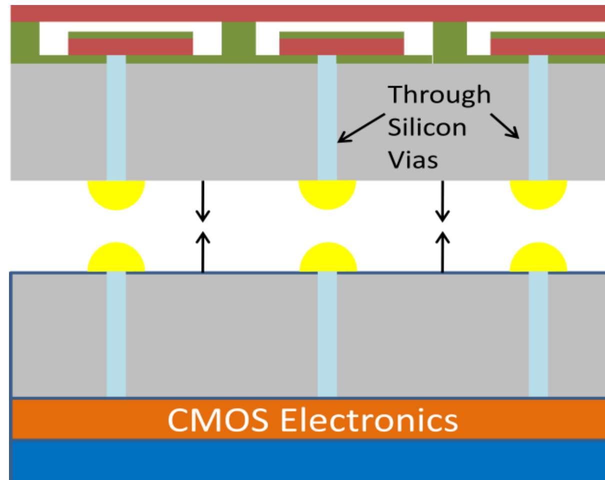


Figure 4.12: Diagram of a flip-chip bonded CMOS-integrated CMUT structure (CMUT structures are shown with a wafer-bonded architecture)

The CMUTs used in the course of this thesis research did not have integrated electronics and as such required alternative connections to the required electronics. Wire bonding is one way to advance from wafer level testing to preliminary packaging. In this method, after the process wafers are diced, thin wires connect the contact pads on the CMUTs to traces on a ceramic package to which the die has been attached. After a CMUT is attached in its package, additional protection can be added to the transducers. Traditional methods used for other MEMS devices like hermetic sealing are not suitable for CMUTs because they require mechanical interaction with the surrounding medium, so biocompatible coatings must be used instead. PDMS is suggested in [69] but other materials such as parylene C can be used. These protective coatings must be carefully chosen so as to not interfere significantly with CMUT operation. The PDMS was found to have no effect on CMUT snapdown voltage or plate characteristics, although it decreased the fractional bandwidth by 21% due to impedance mismatch. The best case scenario could be achieved when using a PDMS material that has good acoustic impedance matching to the medium. For the TOBE CMUTs that were tested and will be shown in chapter 5, a layer of parylene C was deposited as an encapsulation layer, and the resulting effect was only a slight increase in the resonant frequency.

Chapter 5

Top-Orthogonal-to-Bottom-Electrode (TOBE) Arrays[6]

5.1 Introduction to TOBE Arrays

Much of the discussion in the previous section has focused on the fundamental concepts behind CMUT arrays in general, and these concepts both motivate the development of TOBE CMUTs and are integral in understanding how they function. Section 3.3 discussed the benefits of 3-D ultrasound, but also the challenges behind implementing the arrays necessary to perform this type of imaging. Section 4.2 discussed the advantages that CMUTs had over piezoelectric technologies, especially in regards to the ease with which they can be fabricated in complex array designs and integrated with electronic components. The concept of TOBE CMUTs is fundamentally a novel intersection of these two discussions. While there have been other solutions proposed to solve the challenges of 3-D imaging (namely the complicated array structures necessary), none of these solutions were specifically oriented towards taking advantage of CMUTs. In fact, most studies on sparse arrays and synthetic aperture imaging have been done using piezoelectric transducers. Conversely, there has been significant research done to fully exploit the advantages of CMUTs over piezoelectric transducers, but only a small portion of this research has been focused specifically on overcoming the complexity of 3-D imaging. The crossed-electrode array structure combined with the imaging schemes that we have proposed represent a way that the specific properties of CMUTs can be applied directly to solve the aforementioned problems. In this way, TOBE CMUTs potentially represent a quicker way for CMUTs to integrate into the ultrasound technology market, because there is no direct competition from piezoelectric technology.

5.1.1 TOBE Array Structure

In a fully-wired 2-D square array containing N x N elements, a total of N^2 channels are required. A promising design that has been developed by our group and others to address this necessity for a large number of channels is the row-column addressable transducer. In this design, electrodes are connected such that each row and each column of the array is individually addressable, as shown in figure 5.1. In this way, only 2N channels are required to address each element. We have specifically dubbed this design a ‘TOBE’ array, which stands for ‘top-orthogonal – to-bottom-electrode’ array. Similar designs have been proposed by Savoia and Pappalardo et al. in [70] and [71]. However, these designs contain two superimposed mono-dimensional arrays defined using variations in the metallization pattern, where each element can be used to transmit only along one direction, and are not actual 2-D arrays with orthogonal electrodes. The orthogonal electrode design has been studied previously both with piezoelectric

---

[6] *A version of this chapter has been submitted for publication in IEEE Trans. Ultrason. Ferroelectr. Freq. Control*

technology by Yen et al. [72] [73], and with CMUT technology by Logan et al [74] [75]. The imaging schemes proposed in the previous four papers specifically take advantage of the row-column architecture to provide transmit focusing in one direction (lateral or elevational) and receive focusing in the orthogonal direction. In addition to further developing this imaging scheme, we have developed and simulated new imaging schemes which offer better performance and are specifically designed to be used with CMUT technology.

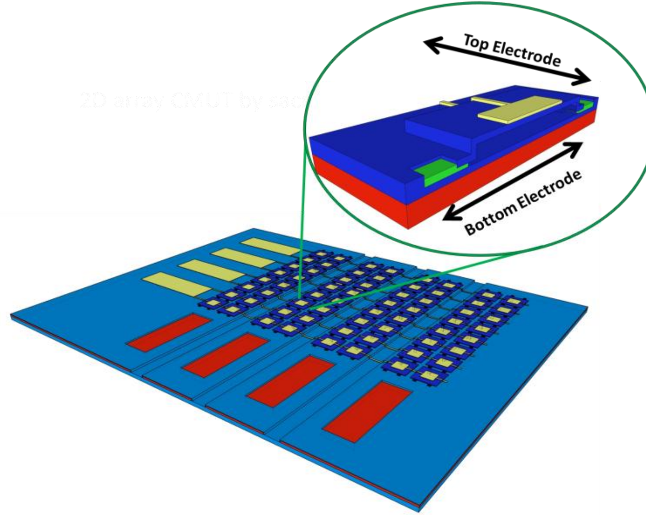


Figure 5.1: 3-D model of our fabricated TOBE array

5.1.2 TOBE Array Fabrication Process

The first TOBE arrays that were designed by our group used a sacrificial release method similar to the one described in section 4.3.1. However, it was modified to allow independent routing of each row or column electrode. The first change is in the substrate; the TOBE process uses an SOI wafer instead of a prime silicon wafer, and the highly doped device layer acts as the bottom electrode. This electrode is patterned using deep reactive ion etching to form the individually addressable column electrodes before any other step is done, after which the rest of the process proceeds in the same manner. The individually addressable row electrodes are defined during the metal electrode patterning step, which also defines the top and bottom electrode bond pads. As previously mentioned, none of the fabricated arrays contained through-wafer vias, and integration with electronics was done through wire-bonding. The TOBE process is, however, conducive to through-wafer via etching in the future if desirable. Because this process uses an SOI wafer, it shares the same advantage that was previously mentioned for the wafer-bonding process. Because the bottom electrode is patterned to be exclusively under the CMUTs, the parasitic capacitance is minimized compared to using a doped wafer as a common electrode. This process was used to successfully fabricate 2-D arrays with elements consisting of 2×2, 3×3, 4×4, and 5×5 individual CMUTs. A scanning electron microscope (SEM) image of one of the 2×2 elements in our array is shown in figure 5.2. A wafer-bonded fabrication process, such as that used by Logan et al. in [74] [75] also

permits the separation of top and bottom electrodes in rows and columns while also providing the improved uniformity and process control of the wafer-bonded process over the sacrificial release process. Our own wafer-bonded process for fabricating TOBE CMUTs was designed and initiated, using the process previously presented in [76]. Completion of this process is still pending.

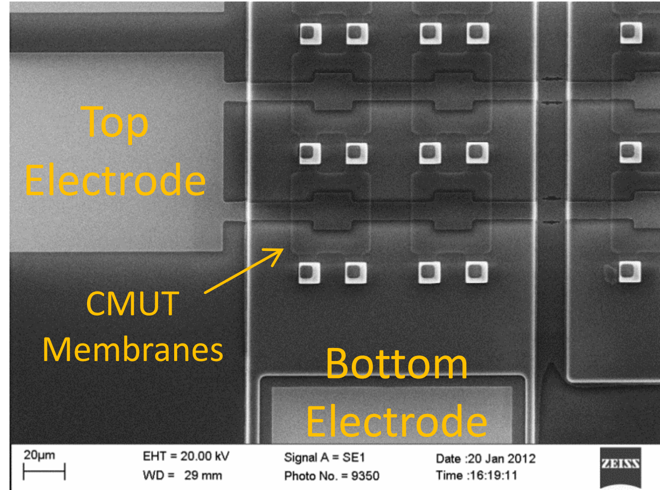


Figure 5.2: SEM image of a single 2 x 2 element in one of our TOBE arrays

5.2 Imaging using TOBE Arrays

5.2.1 Imaging Scheme 1

Although the imaging schemes were initially referred to only by number, a descriptive name that would provide a good idea of the principles behind this imaging scheme could be: "Orthogonal one-way focused synthetic aperture imaging". This is the method that has previously been proposed by Yen et al [72] [73] for piezo-electrode crossed-electrode arrays and Logan et al [74] [75] for CMUTs. In this method, which is shown in figure 5.3, transmit pulses and bias voltages are applied on the row electrodes while the orthogonal column electrodes are held at ground. After each transmit event, the row electrodes are then clamped at a given voltage, and signals are received by the column electrodes. Unlike with piezoelectrics, these columns can be biased to improve sensitivity. As such, this method is similar to the synthetic aperture methods described in section 3.4, but the signals are received orthogonally to the direction in which they were transmitted. Both transmitting and receiving can be done using one or multiple rows/columns. For example, the signals could be received with dynamic receive focusing if the proper delays are applied to all the columns. After all the signals are received, retroactive focusing can be done using the same linear array delay formulas described for synthetic aperture imaging in section 3.4. Logan et al use multiple elements on transmit to form a cylindrically-focused beam rather than the column-by-column transmit scheme shown in figure 5.3. This has the disadvantage of poor focusing away from the cylindrical focal zone. The exact imaging method that was simulated will be described in more detail in the section

demonstrating the simulations. Using scheme 1, a 3-D image may be formed using only N transmit events, and with N transmit channels and N-receive channels. Thus, relatively fast imaging can be achieved using much fewer channels than a fully-wired array. The disadvantage of scheme 1 is that only one-way x-focusing and one-way y-focusing can be implemented, hence the contrast-to-noise ratio will be poor due to off-axis sidelobe-induced clutter.

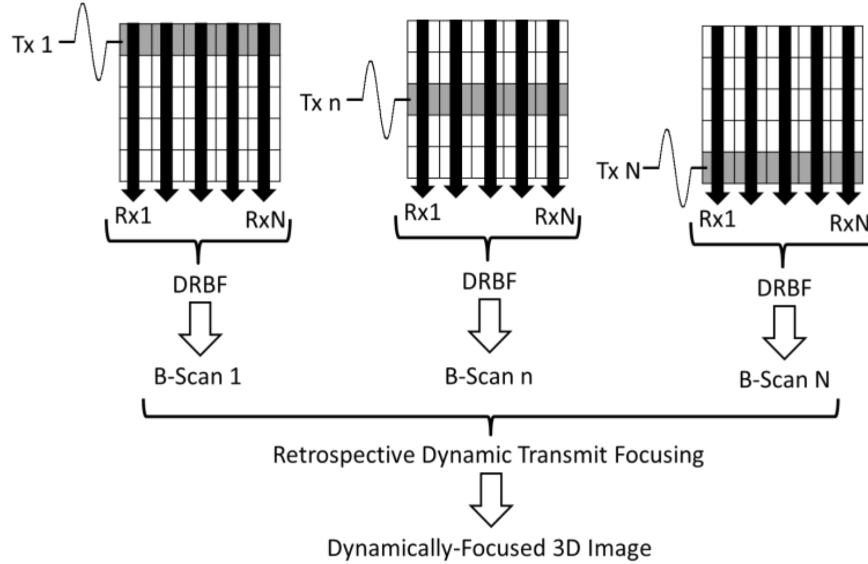


Figure 5.3: Imaging Scheme 1: Orthogonal one-way focused synthetic aperture imaging

5.2.2 Imaging Scheme 2

We have developed a new interfacing scheme that can be known as orthogonal two-way actively focused synthetic aperture imaging. This scheme takes advantage of the nonlinear transmit and receive response of CMUTs to applied bias voltage, and, unlike scheme 1, cannot be implemented using conventional piezoelectric technology. By exploiting the ability of CMUTs to display higher dynamic displacements when oscillating with a bias voltage near its snapdown voltage, this method can achieve effectively single element control. If a transmit pulse is applied along a row, certain elements in that row can achieve much higher displacements than the others, simply by applying a bias voltage to their corresponding columns and grounding the rest of the columns. Similarly, a membrane which is biased during receive mode will be more effective, producing a current with a higher amplitude than one which is not biased, resulting in selective element receiving along a row. By allowing individual element control along any row, there is significantly more flexibility in terms of possible imaging methods that can be implemented. For example, it could be possible to implement fully sampled 2-D array synthetic transmit aperture/synthetic receive aperture or combinations thereof, as shown previously in [24] [37].

The initial imaging method we propose (depicted in figure 5.4) is as such: for a single biased column, several transmit/receive events (scanlines) can be performed with active transmit focusing and dynamic receive focusing in the

elevational direction to construct a B-scan. After obtaining a B-scan for every column, retroactive dynamic transmit focusing can be performed in the lateral direction, again using the same beamforming methods covered in section 3.4 for synthetic aperture imaging. Overall, this imaging method can achieve two-way focusing in the elevational direction, and one-way focusing in the lateral direction. This results in a higher quality image compared to scheme 1, but it comes at the expense of significantly more transmit events, requiring $N \times S_Y$ events for an $N \times N$ array, thus resulting in slower imaging. However, the number of transmit events is still likely less than the $S_X \times S_Y$ events required for sector/sector pyramidal scanning, while maintaining the advantage of only requiring N transmit channels and N receive channels.

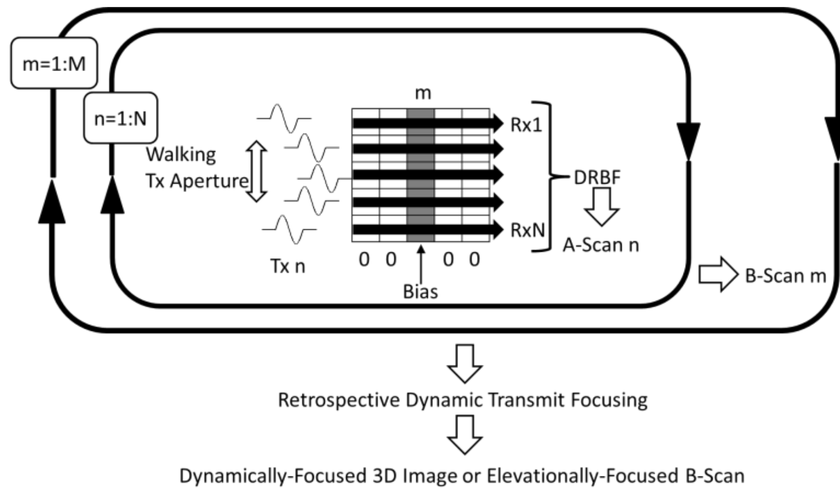


Figure 5.4: Imaging Scheme 2: Orthogonal two-way actively focused synthetic aperture imaging

Although this imaging method has the potential to use fewer transmit events than traditional scanning methods, another consideration that must be taken into account is bias voltage switching time. Because the bias voltage of a particular column needs to be switched on after the previous column is used for imaging, and off before the next column is used, if the switching time is too long, the frame rate advantage could be lost. For example, at a 20mm imaging depth, for a 64×64 element 2-D array with 100 scanlines, the bias voltage switching time would need to be less than 731µs to obtain a faster frame rate than a pyramidal scan using the same size 2-D fully-wired array and same number of scan lines in both directions. Long data acquisition times could cause motion artefacts in the case of imaging moving tissue, so reducing these times is highly desirable. There are commercially available high-voltage switches with ~µs switching time which could potentially achieve fast enough switching rates. Additionally, previous work on collapse-snapback operation modes has shown that membrane response to applied bias voltage can be fast enough to keep up with fast switching times [77].

5.2.3 Imaging Scheme 3

The imaging method proposed above is very similar to the monostatic classical synthetic aperture approach. The main difference is that, instead of having one element transmit and receive at a time, one column is transmitting and receiving at the same time, and they are actively focusing in the elevational direction. While higher quality images are obtainable with this method as compared to scheme 1, it still suffers from the same problem as linear array-based SAI methods – that of low SNR due to weaker transmit pulses. Additionally, because only one element at a time is effectively transmitting (due to focusing delays), parasitic capacitance from interconnects could create even more problems, further degrading SNR. Finally, it is desirable to have an imaging scheme which can provide two-way focusing in both the lateral and orthogonal direction. To address all these problems, we propose a novel coded aperture pulse sequence that has the potential to significantly improve signal-to-noise ratio and resolution at the cost of additional transmit events.

As has been explained previously, numerous variations on synthetic aperture imaging have been proposed which use multiple elements upon transmission and reception to improve SNR. Using multiple sets of receive elements for each set of transmitting elements can also provide better resolution by capturing more of the reflected signal. These methods can be applied to 2-D arrays as well, for example by using multiple columns instead just a single column for transmission and reception. For TOBE arrays, this can be done by applying a high bias voltage to more than one column. The difficulty in extending this concept specifically to TOBE arrays is that the individual response from a specific element cannot be isolated; the received signal will contain the response from all the biased elements in a particular row or column. To be able to isolate the signals in such a way that they can be beamformed using conventional SAI methods, some non-trivial mathematical analysis is required. Appendix 1 contains the conference proceedings which details this analysis. The imaging method can be summarized by the process shown in figure 5.5. The columns active during each transmit event are determined by binary values in the rows of a particular type of matrix called an S-matrix. As shown in appendix 1, using this matrix results in optimal SNR and performance. We use this S-matrix aperture encoding sequence in simulations to demonstrate that the highest quality images can potentially be achieved with a TOBE array

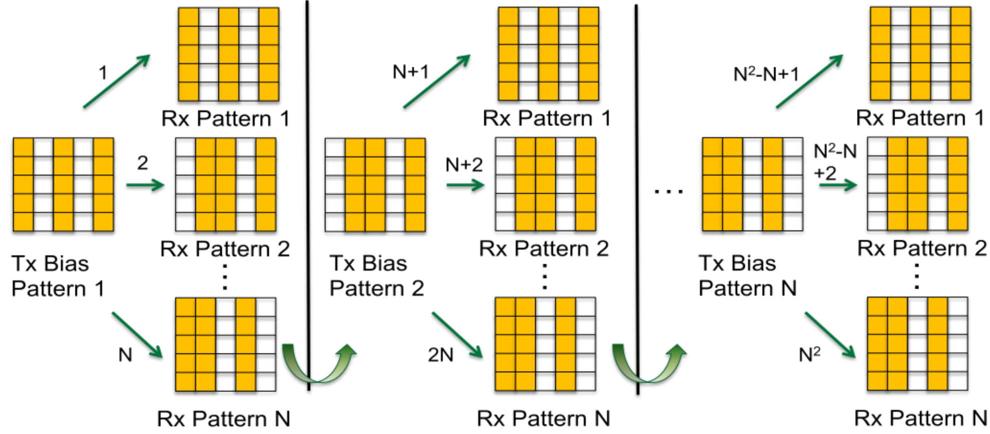
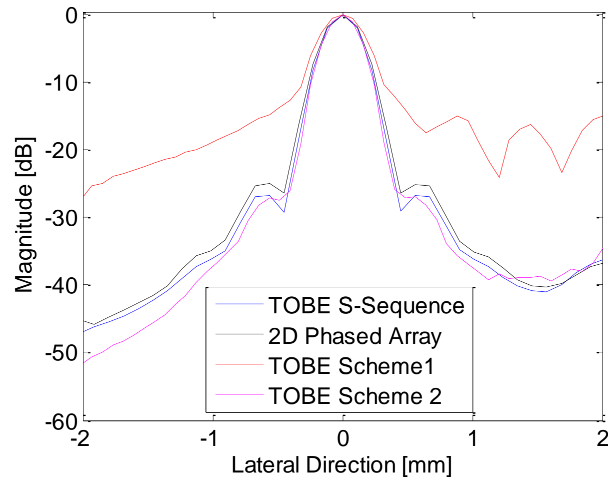


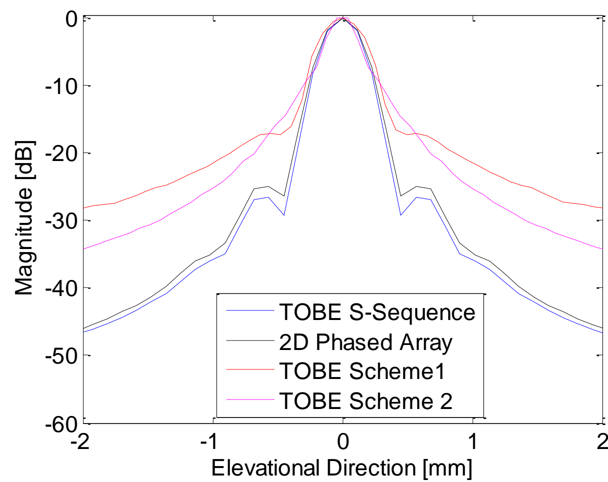
Figure 5.5: Visual representation of the S-sequence aperture encoded imaging method

5.2.4 Imaging Simulations

To model the relative imaging performance of the three different imaging schemes, the ultrasound simulation software Field II [78] [79] and the beamforming toolbox developed by Jensen and Nikolov [80] were used. For schemes 1 and 2, a 192-element by 192 element TOBE array was used. The element widths and kerfs were 0.87 λ and 0.087 λ respectively in both x- and y-directions. A walking aperture rectilinear scanning approach with zero-steering angle was used with 64 active elements. For scheme 3, a 63 x 63 element TOBE array with the same element and element spacing parameters was used with steering. For comparison purposes, a fully-wired 2-D phased array was also simulated. The maximum amplitude projection C-scan image of two points located at an imaging depth of 104-wavelengths from the array surface after was found using all four methods, and the cross sections of each C-scan were taken in both elevational and lateral directions to determine obtainable resolution. These results are shown in figure 5.6. It should be noted that for the scheme 1 simulation, transmitting was done along individual rows and dynamic receive focusing was done along columns. For scheme 2, the bias voltages were applied on the rows, meaning each transmit/receive event was focused in the lateral direction (the description of the imaging scheme in the previous section referred to transmitting/receiving along columns, but the process is analogous).



a)



b)

Figure 5.6: Cross-Range-Maximum-Amplitude point-spread function plots comparing three different imaging schemes with a fully-wired 2-D phased array in: a) the lateral direction and b) the elevational direction

From figures 5.6 a) and b), we can see the relative performance of each imaging method. The stand-out imaging scheme is the S-sequence aperture encoded method (scheme 3), which matches the resolution of the fully-wired 2-D phased array in both directions. In the lateral direction, where scheme 2 has two-way focusing, it can achieve nearly identical resolution to a fully-wired array. However, as expected, in the elevational direction, where the only focus is retroactive transmit focusing, the performance suffers compared to a fully-wired array. Scheme 1, which has only one-way focusing in both directions, displays considerably worse resolution than all three other schemes in the lateral direction, and even in the elevational direction, has higher sidelobes than scheme 2.

While resolution is one way to determine imaging quality, another important metric is SNR. The same simulations displayed above were repeated, but a constant noise level was added all the simulations to simulate realistic

system noise. In addition, a simulation was done where instead of using S-matrices to determine the active Tx/Rx columns, identity matrices were used instead. Figure 5.7 shows side-by-side comparisons of each imaging method, and the relative SNRs are given in table 5.1.

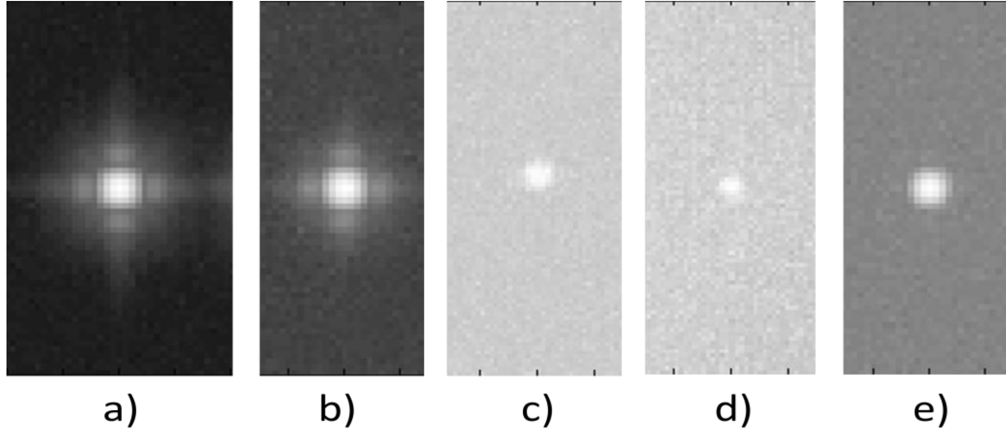


Figure 5.7: Noise-added C-scan point spread functions of: a) S-sequence aperture encoding scheme; b) 2-D phased array; c) Scheme 1; d) Scheme 2; e) Identity matrix aperture encoding scheme

Table 5.1: SNR Comparisons Between Various 3-D Imaging Schemes

Imaging Scheme	SNR
TOBE S-Matrix Imaging Scheme	+48.3 dB
2-D Phased Array	+40.3 dB
Identity Weighting Matrix Scheme	+26.5 dB
Previously Proposed TOBE Scheme 1	+11.1 dB
Previously Proposed TOBE Scheme 2	+9.9 dB

As can be seen from these results, the SNR of the S-matrix encoded scheme is vastly improved compared to schemes one and two. The overall simulation comparison demonstrates the ability for TOBE CMUTs to provide varying imaging schemes with different abilities based on the necessities of the application at hand. Because scheme 1 can achieve a high frame rate, requiring only N transmit events to obtain an image, at the cost of having only one-way focusing in both lateral and elevational directions, it could be used in situations where real-time imaging is necessary, but high resolution is not, such as imaging of larger moving structures. Scheme 2 could still achieve faster imaging than fully-wired 2-D arrays with better resolution in one direction than scheme 1, and

could therefore be used in situations where resolution is more of a necessity. Finally, scheme 3, with its potential to match the resolution of fully-wired 2-D arrays at the cost of more transmit events, could be more suitable for industrial applications such as NDT, where imaging speed is not as critical.

5.3 Feasibility Test Results

The imaging schemes presented above show potential for the use of TOBE CMUTs in 3-D imaging in a number of different capacities. However, the effectiveness of the presented imaging schemes depends entirely on the ability to obtain single/dominant-element transmit and receive actuation with the TOBE array. For the arrays to show any promise at all, the ability to control a single element at a time using the row-column addressing method must be shown. A major concern for all ultrasound arrays is crosstalk between transducer elements. The major source of mechanical crosstalk is the Stoneley and Lamb waves introduced in section 2.1.1. Electrical crosstalk can be caused by capacitive coupling between elements. Crosstalk causes elements in an array other than the desired active element to transmit signals, and reduces image quality. It can be especially important to account for in 2-D arrays, where the elements must be very close together. Additionally, for the TOBE array structure, there is potential for electrical crosstalk to be even more pronounced, since all elements are interconnected through the row-column architecture. As such, it is important to verify how crosstalk affects single-element operation before validating the potential of the previously mentioned imaging schemes.

5.3.1 RF Actuation Vibrometer Tests in Air Operation

To perform preliminary validation of the proposed imaging schemes, a laser vibrometer system was used to measure membrane displacement in air for a sample 32×34 array with each element containing 9 separate membranes in a 3×3 arrangement. First, however, the array was tested in a configuration similar to scheme 1 to verify whether any electrical crosstalk caused by capacitive coupling between elements would significantly increase crosstalk of un-actuated columns compared to an array where the columns were not interconnected. The test setup is shown in figure 5.8. An 8.55 MHz 2Vpp sine burst with 45 cycles and 10μs trigger interval was applied along the first column which was biased with 120V. All the other electrodes were shorted together and grounded except for the bottom row, which was floating. The highest displacement value in each 3×3 element was recorded for a 4×4 subsection of the array, and the recorded data is shown in figure 5.9. It should be noted that the displacement of the elements in row 1 is not included since this row was not grounded as would normally be done in scheme 1 imaging. Even taking the anomalous low displacement of row 5 in column 1, the worst-case cross talk is approximately -13dB (with the average cross talk being closer to -29dB). By comparison, when an element in one of our linear arrays on the same wafer (element 1 in figure 5.10 was actuated, the crosstalk in elements 2 and 3 was as high as -12.5dB and -9.6dB respectively. This suggests that, compared to more electrically isolated linear array elements, un-actuated rows/columns in the TOBE array have little or no significant additional electrical

cross-talk due to capacitive coupling between the rows/columns. Were steps taken during fabrication to minimize mechanical crosstalk due to Lamb waves, it is probable that the crosstalk could decrease even further. Additionally, using a wafer-bonded fabrication method would improve the uniformity between elements and lead to less variation in crosstalk values.

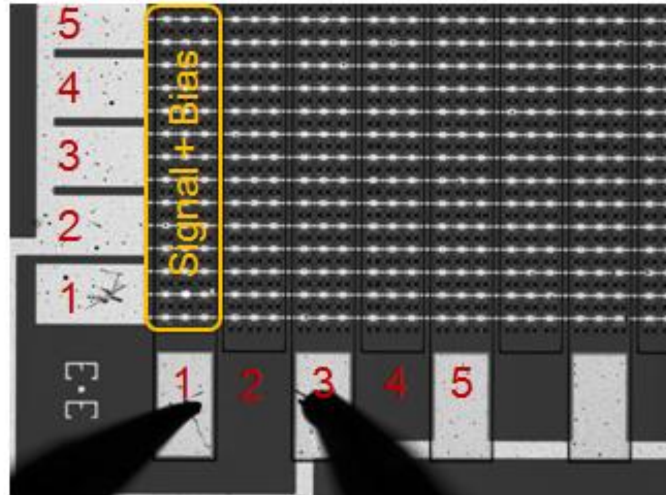


Figure 5.8: Scheme 1 test set-up: all columns and rows are grounded except for column 1 and row 1 (which is floating)

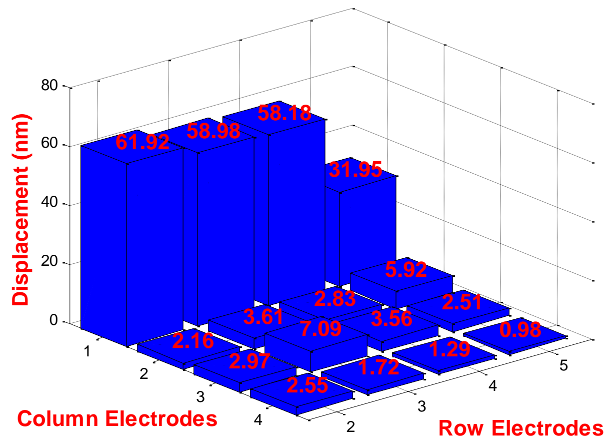


Figure 5.9: Dynamic displacement vibrometer measurements of a 4 x 4 section of a TOBE CMUT array operating in scheme 1

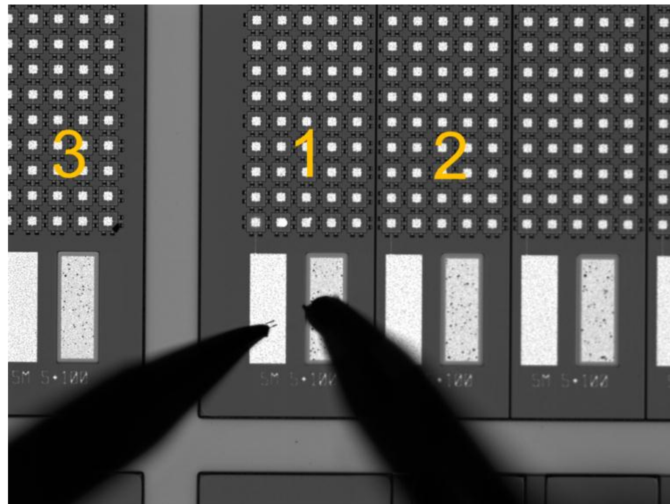


Figure 5.10 Linear array test set-up for comparing crosstalk

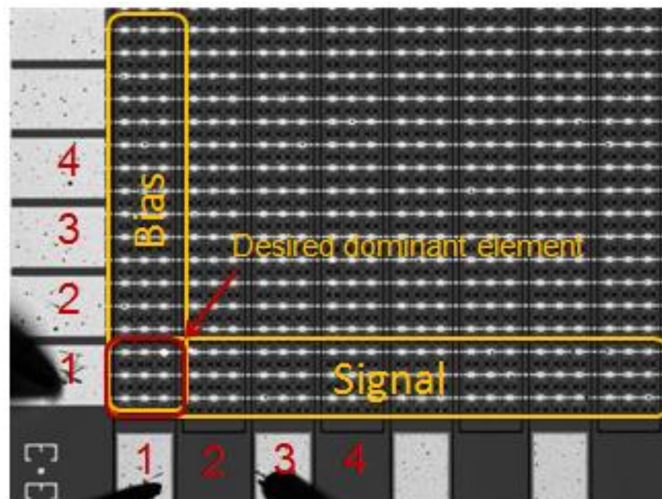
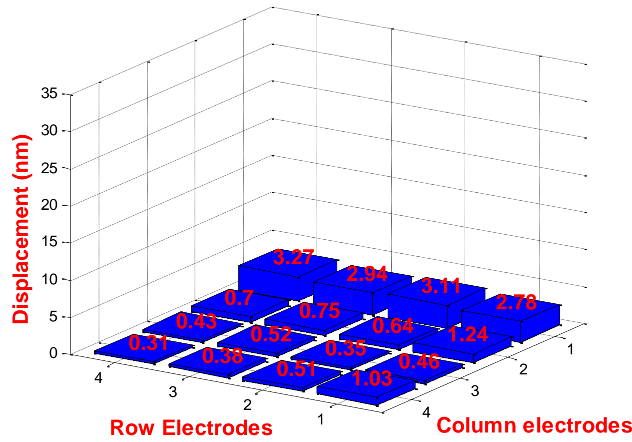


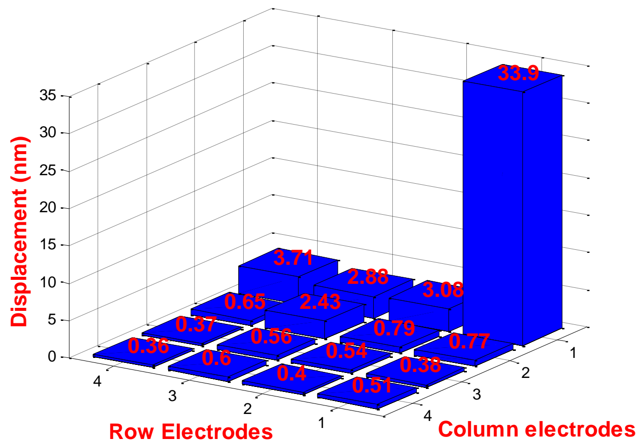
Figure 5.11: Single-element actuation test set-up: all columns and rows are grounded except for column 1, which is biased, and row 1, which has signal applied

Afterwards, the array was connected as shown in figure 5.11 to perform preliminary tests on the ability to achieve dominant element actuation – specifically dominant element transmit. An 8.58MHz 2Vpp square pulse burst with 45 cycles and 10µs trigger interval was applied along row 1. All other rows and columns were grounded other than column 1. Figure 5.12 shows the maximum displacement in each 3x3 element for two different tests. In figure 5.12 a), the first column was also grounded. In figure 5.12 b), the first column had a 70V bias applied to it. It can be seen that increasing the bias voltage on the first column increased the displacement of the desired element by over 12 times, while having a negligible effect on the displacement of the other elements. These results show that single element actuation is easily achievable in air operation, aided by the fact that the biased membrane has a lower resonant frequency than the unbiased membranes (due to the spring softening effect mentioned in section 4.1).

Because of the very small bandwidth of CMUTs in air operation, the increased displacement of the biased element at a lower frequency does not produce any measureable cross talk. Further selectivity to the biased element could be achieved by subtracting the measured response of figure 5.12 a) from that of figure 5.12 b) to remove any response from the other elements and to yield even better single element control during transmit mode.



a)



b)

Figure 5.12: Dynamic displacement vibrometer measurements of a 4 x 4 section of a TOBE CMUT array operating in scheme 2 with: a) 0V bias on column 1; b) 70V bias on column 1

Operating the active elements in collapse mode was also investigated as a means to further reduce crosstalk in inactive elements. As previously mentioned, collapsed membranes have a much higher resonant frequency than pre-collapsed ones. Actuating these collapsed membranes at their resonant frequency thus induces a negligible vibration on adjacent membranes which are un-collapsed,

and hence reduces crosstalk. Tests were done comparing the dynamic displacement in air of collapsed membranes with that of un-collapsed ones. The results are shown in table 5.2. A clear improvement over the crosstalk observed in pre-collapse operation is visible. However, through the course of collapse mode operation, significant charging was observed which resulted in the continual increase in snap-down voltage over time. After several minutes of operation, some membranes would snap back up despite being held at constant bias voltage. Although stored charge can be eliminated by reversing the polarity of the applied voltage, requiring this additional step for efficient CMUT operation causes charging to remain a significant drawback to collapse mode operation.

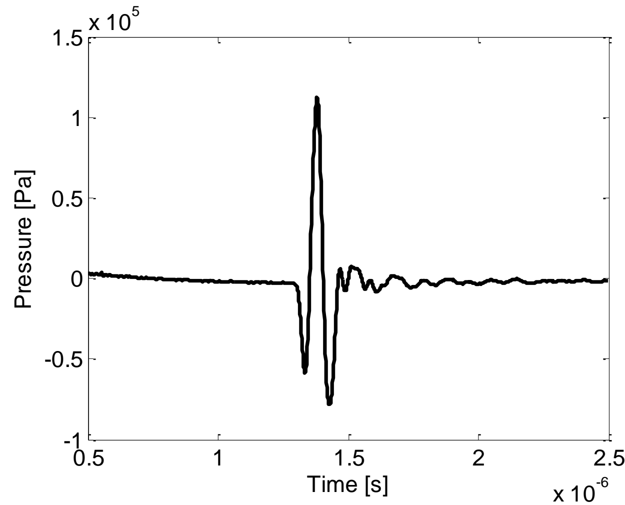
Table 2: Collapse-mode TOBE CMUT Operation and Cross-talk Comparisons

Operating Mode	**Actuated Membrane Displacement**	**Un-actuated Membrane Displacement**	**Improvement over Pre-collapse Mode**
Continuous Sine Wave	22.5 nm	0.3 nm	3.75 x
Pulsed Sine Wave	1.2 nm	0.045 nm	1.33 x

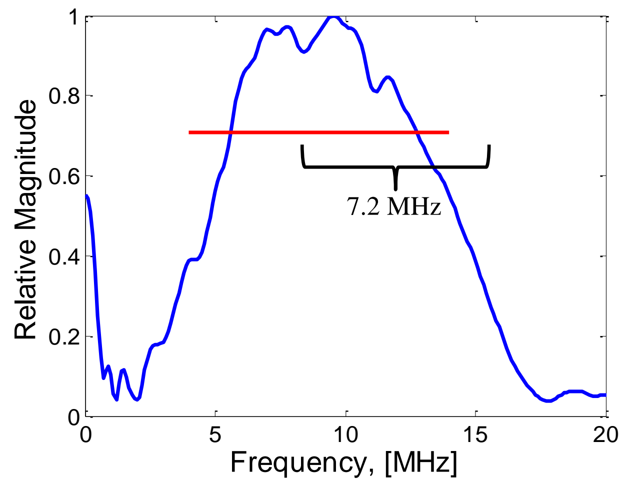
5.3.2 Acoustic Immersion Transmit Testing

After being tested in air, pressure measurements were obtained from the TOBE CMUTs while immersed in vegetable oil to validate dominant-element transmission in immersion. A 7mm x 7mm die was wire bonded to a CFP80 ceramic package, and then coated with parylene C for improved sealing and electrical isolation. The package was then mounted on a custom circuit board and tank assembly and the tank was filled with a small amount of vegetable oil. A hydrophone with a +20dB pre-amplifier was connected to an oscilloscope to measure the generated pressure waves. The array tested was a TOBE array with elements containing 16 membranes in a 4x4 arrangement. All of the columns were grounded except for two, which were connected to a DC power supply and biased with 70V. All of the rows were grounded except for two, which were driven using a function generator. The result was an active area corresponding to a 2 x 2 element section measuring 456μm × 476μm, as shown by the yellow lines in figure 5.14 b). The applied signal was a 1Vpp, 40ns square pulse with a 10μs trigger interval that was first passed through a 35 dB gain power amplifier. To determine the pressure field, the hydrophone was first moved to the on-axis point at which the recorded hydrophone signal was greatest. This signal and its corresponding Fourier transform are shown in figure 5.13. The hydrophone was then moved in both the x- and y-directions to obtain pressure field measurements in these directions. The hydrophone was located approximately 2mm from the array to minimize measurements from adjacent elements. The pressure signals recorded in the z = 2mm plane effectively formed a receiver array and were used

to form an image of source pressures using delay and sum beamforming with receive-only delays, in a way similar to photoacoustic beamforming, as explained further in [81]. This visual representation of the pressure field is shown in figures 5.14 a) and b). It was found that the pressure bottomed out at -18dB past 700μm in the x-direction, while in the y-direction, the pressure continued decreasing until no measurable pressure could be obtained with the hydrophone. The residual pressure in the x-direction is due to the driving signal that actuates every element along one entire row.

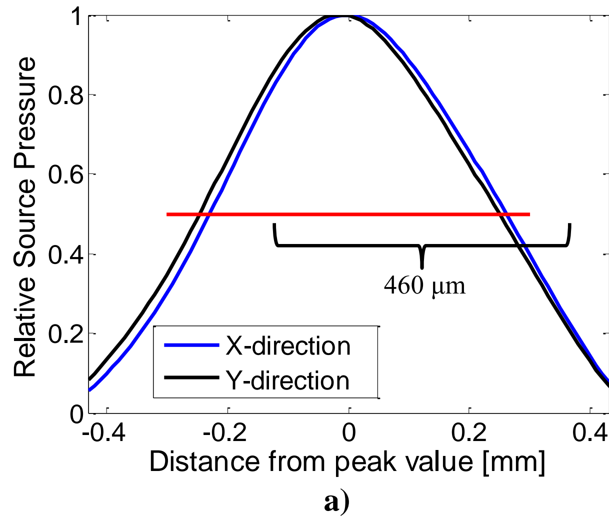


a)

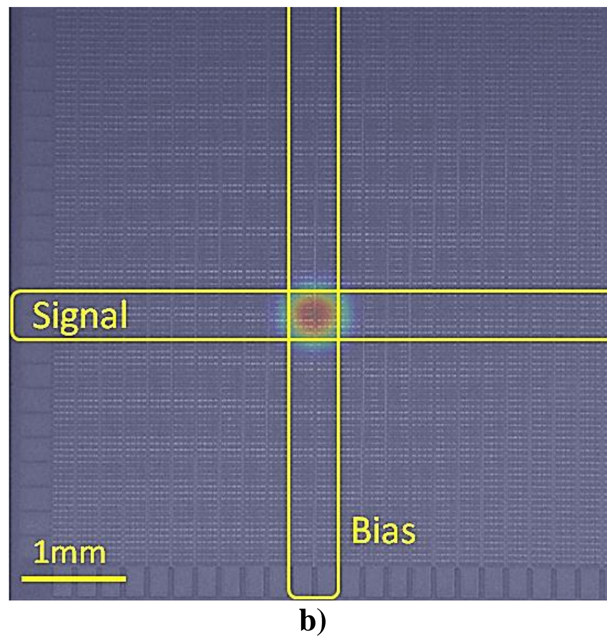


b)

Figure 5.13: a) Maximum recorded pressure signal from TOBE array; b) Fourier transform of pressure signal



a)



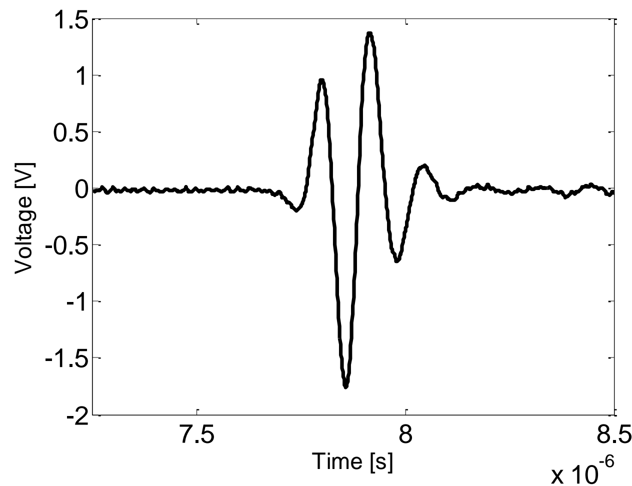
b)

Figure 5.14: a)Near-field x- and y-direction pressure variations of TOBE array. b) Pressure field map of entire TOBE array superimposed on SEM image of TOBE CMUT die

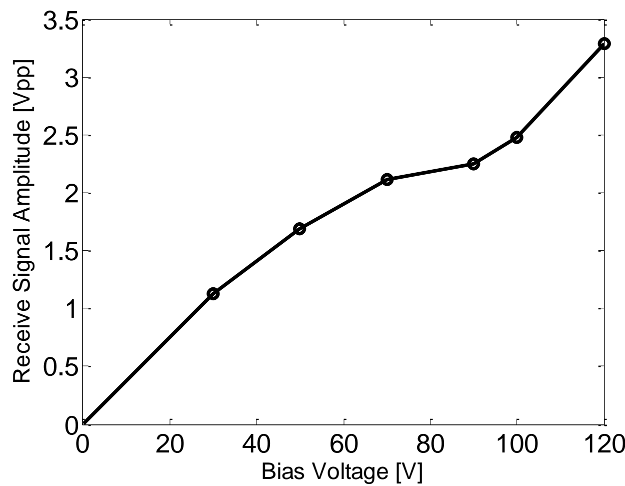
5.3.3 Acoustic Immersion Receive Testing

The performance of the TOBE CMUTs in receive mode was also tested in immersion. The same custom tank and packaged TOBE array assembly described in the previous section was used. The same two columns were connected to the DC power supply, while the two rows initially used for transmitting were connected to an oscilloscope through an amplifier. First, an unfocused 10MHz transducer was connected to the function generator and immersed in the oil above the TOBE array, so as to measure the sensitivity and overall receiving capabilities of the array. The 10MHz transducer was chosen because it was the available transducer which most closely matched the measured transmit center frequency.

The signal provided to the external transducer was a single sinusoidal pulse with frequency of 8.4MHz (the frequency at which the measured signal was found to be greatest). To test the sensitivity, the active columns were biased at 100V, and the amplitude of the external transducer signal was decreased until the received signal was no longer observable. It was found that the minimum observable signal corresponded to a pressure of 300 ± 30 Pa, or a noise equivalent pressure (NEP) of ~0.1Pa/$\sqrt{Hz}$. The received signal, amplified by 63dB, is shown in figure 5.15 a). The effect of bias voltage on receive sensitivity was also tested using the unfocused transducer. While applying the same input signal with a 10V amplitude to the external transducer (corresponding to 10.6kPa), the bias voltage on the two active columns was varied, and the output signal was measured. The results are shown in figure 5.15 b). At a 120V bias, the NEP further improved to 83mPa/$\sqrt{Hz}$.



a)



b)

Figure 5.15: a) Maximum recorded received signal from TOBE array at 100V bias; b) Change in maximum received signal amplitude due to variations in bias applied bias voltage

To test the contribution of each individual receiving element to the total signal received, the transducer was switched out for a focused 10MHz transducer. This transducer was held at a distance above the TOBE array equal to its focal length (1.9cm). The FWHM of the focused pressure profile at this point corresponded to 683µm. This pressure profile would thus extend approximately 110µm past the active area assuming that the transducer was centered in the middle of the active area. From the on-axis point where the received signal amplitude was highest, the transducer was then moved in the x- and y-directions and the received signal was measured. This was done while the columns were held at constant bias voltage. The results obtained for a bias voltage of 80V are shown in figure 5.16. It can be seen that ~1.6mm away from the center, the signal was only ~5% of the maximum signal. At a bias voltage of 100V, this decreased further to 4.7%.

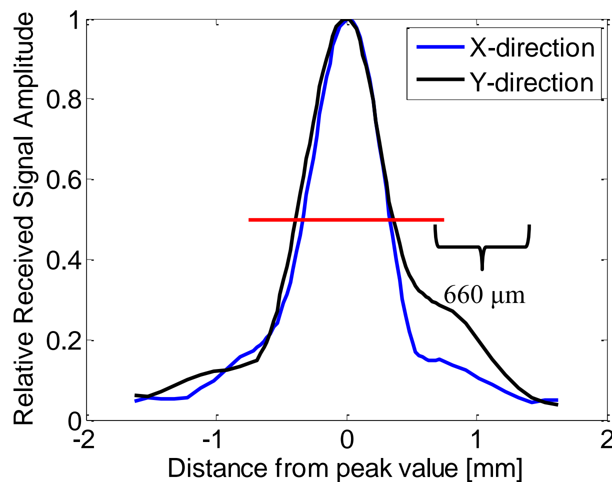


Figure 5.16: Received signal variations of TOBE array due to translation of a focused transducer away from active area

<u>5.4 Interface Electronics</u>

In the previous experiments, the electronics consisted of a simple interface board with only 20 available channels that could be connected to only specific pins in a mounted package. While this setup is suitable for feasibility experiments, it is not suitable for more detailed imaging validations. As such, work has begun on electronics that will be capable of interfacing our developed CMUTs to a Verasonics imaging system. While completion of this system may not occur before the completion on this thesis, a description of the system will be given to summarize the design. This design was specifically made to be able to interface with the TOBE CMUTs and allow for the implementation of the previously described imaging schemes.

The Verasonics imaging system has the capability to interface with 128 channels. In the current design of the interface electronics, 80 channels are used so as to easily coincide with the 80-pin ceramic packages already available. The transducer interface board consists of an area where these connectors are mounted, and all 80 channels are connected to 10 pre-amplifiers from Maxim Integrated (eight channels per amplifier). These amplifiers, specifically designed

for use in CMUTs, and their placement are integral to the design. Because of the low capacitance of an individual TOBE CMUT element, removing the amplifier, or adding significant wiring in between the element and the amplifier would cause the associated parasitic capacitance to swamp the received signal from the CMUTs and result in low SNR. The required supply voltages for these amplifiers are provided by a separate power-supply board designed to provide low-noise DC power sources. From the amplifiers, the channels are then connected to the Verasonics system through a specially designed interface which uses a 260 pin ZIF connector. The novelty of these electronics is in the requirement for each channel to have an individually addressable bias voltage, so as to allow for scheme 2 and 3 imaging. Previously designed ultrasound interface electronics would have been able to use just one DC bias source for all the CMUTs, since individual element control was done by switching Tx/Rx channels, not bias voltages. The need for bias voltage switching necessitated the addition of a separate bias voltage control board. This board contains a high voltage latched shift register, with power supplies provided by the same power supply board, and logic signals provided by a microcontroller connected to the Verasonics host PC for easy software control of logic signals. With this design, the specific configuration of bias voltages on each channel could be provided using 80 serial bits, then provided to the CMUT interface in parallel. Each high voltage output would be connected to a separate transducer I/O channel in between the pre-amp and the ceramic package through a bias tee. With these electronics, it will become possible to experimentally verify the imaging capabilities of not only the TOBE CMUT arrays, but any other arrays designed in the future.

5.5 Discussion of Results

Overall, the results obtained in section 5.3 show potential for the imaging schemes we have envisioned for the TOBE arrays. The elements in the tested array displayed a center frequency of 9.55MHz with a 75% (7.2MHz) -3dB fractional transmit bandwidth. A 40 ns pulse was chosen to increase the strength of the transmitted signal, but a shorter pulse could be chosen as well, which would increase the transmit bandwidth. The receive signal was likely band-limited by the external transducer, which did not have the same center frequency as our TOBE array, so the frequency response figure is not included, since its accuracy could be limited.

While, to our knowledge, there has not been a comprehensive experimental analysis and characterization of crosstalk in 2-D CMUT arrays (much more extensive research has been done on linear arrays), other publications on 2-D ultrasound arrays have reported varying values. Lee et al [82] and Dausch et al [83] report typical crosstalk values of -30dB and -45dB respectively for their 2-D piezoelectric arrays. For CMUTs, Bayram et al [84] found crosstalk to average -23.2 dB for pre-collapse operation and -34.4dB for 1D-CMUT arrays in collapse operation, averaged across a distance of 4.5mm from the active element. In our single-element immersion transmit tests reported in section 5.3.2, the crosstalk between the active element and the other elements in the same row (which were all being driven by the same signal) when the active element was biased at 70V

bottomed out at -18dB past 700μm, while the pressure signal from elements in other directions was not even observable. In vibrometer testing, the crosstalk of most other elements was less than -30dB, and given that the bottomed-out immersion crosstalk along the transmitting row is similar to the maximum air-coupled crosstalk, we can infer that immersion crosstalk across most of the array would also be similar to -30dB. This is comparable to the previously reported values and less than the average column-to-column cross talk obtained.

Since it was observed that increasing the bias voltage resulted in a much greater change in displacement for the active element compared to adjacent elements, it would be possible to decrease the crosstalk even more by further increasing the bias voltage, especially in air, since the spring softening effect would further decrease the resonance frequency of the active element relative to elements in other columns. As mentioned previously, it would also be possible to theoretically improve an image obtained in air by subtracting the zero-bias response from the high-bias response to effectively remove any response from other elements (since their displacement would remain unchanged). This method would not be as feasible in immersion operation because the increased bandwidth of the membranes and the increase in mechanical crosstalk in immersion result in more significant displacement changes for adjacent elements.

For the acoustic immersion transmit tests, the full-width-half-maximum (FWHM) of the pressure field was approximately 460μm, very similar to the dimensions of the active area. The pressure map in figure 5.14 confirms this, strongly suggesting that most of the transmitted signal is being emitted from the active area. For our tests, the active area was a 2×2 element section because of a limitation in the number of channels on our test board. The active area could be even smaller were only one element active, and for imaging tests using the new custom interface electronics, this will be possible.

Our receive tests are also conducive towards the potential of dominant element receiving. The strong dependence of received signal strength on applied bias voltage is a strong indicator that only the biased elements have any significant contribution to the total received signal. The focused transducer tests confirm that this contribution is largely limited to the 2×2 active area. Although the FWHM of the received signal variations is larger than the size of our active area, this can be attributed to the fact that the FWHM of the external transducer's focal pressure field is larger than our active area, meaning that even when the transducer is further than ~230μm along the x- or y-directions from the center, it is still appreciably actuating the active area. In fact, the transducer FWHM is almost identical to the FWHM obtained in figure 5.16, further supporting the idea of single- or dominant-element receiving. As with transmit mode, the contribution from inactive elements could be even further reduced by further increasing bias voltage.

Our arrays were also not optimized to minimize mechanical crosstalk, so it is feasible that crosstalk could be decreased further by implementing methods to reduce crosstalk during fabrication. Trench isolation and wafer thinning are two examples that have been proposed to reduce Lamb-wave coupling [49]. The observed sensitivity is not as high as some of the CMUTs reported in literature

with a noise-equivalent pressure of ~1mPa/, but it is likely that it could be improved by using more optimized electronics. In the future, devices could be fabricated using a wafer-bonding process to improve device quality relative to the sacrificial release method. Overall, test results demonstrated that both dominant-element transmitting and receiving are feasible using our proposed interfacing scheme. This would therefore allow for 3-D ultrasound imaging using only N transmit-receive channels and N bias-control channels and result in drastically decreasing the cost and complexity of 3-D imaging.

Chapter 6

Conclusion

6.1 Summary

The purpose of this M.Sc thesis was to present the work done throughout the course of the degree on TOBE CMUT arrays. The overall result of the work done was the design, simulation, characterization, and understanding of these 2-D CMUT arrays and their respective imaging schemes. This necessitated a thorough knowledge of the fundamental concepts of acoustics, ultrasound imaging, and CMUT operation and fabrication. A summary of the work done and its related fundamental concepts is presented, as well as possibilities for future work that could be done using the TOBE arrays.

Ultrasound technology is valuable as an imaging system due to its relative low cost compared to other imaging systems such as MRI and CT, its use of non-ionizing radiation, and its simplicity and effectiveness. A significant gap in this technology lies in the ability to easily obtain 3-D images. The associated difficulties stem from the basics of ultrasound imaging and acoustics and how they relate to imaging performance. High quality images are characterized by high resolution, which is directly proportional to the size of the array (and hence the number of separate transducer elements). The periodicity of the element arrangement, however, results in grating lobes in the image if the elements do not satisfy specific spacing requirements. Focusing and steering of an ultrasound beam can be done electronically by applying delays to each separate element. Focusing further increases the resolution of the image, whereas steering can increase the size of the imaged area without increasing the size of the array by enabling off-axis focusing. However, beam steering can further enhance the presence of these grating lobes. As such, to avoid grating lobes, element pitch must be less than half the wavelength of the transmitted ultrasound. For 3-D imaging, this requirement is present in both the lateral (x-) and elevational (y-) directions, and thus the combined requirements of large arrays for high-resolution imaging, and closely spaced elements for grating-lobe free imaging, result in the need for a large number of densely packed elements in a two-dimensional architecture. Connecting every element to a separate electronic channel is very difficult and cost-intensive, and for piezoelectric ultrasound transducers, which constitute the majority of commercially available transducers, the fabrication process becomes more difficult. Various methods have been proposed to address these problems, such as sparse array imaging, which removes elements out of the array to eliminate or alter the element periodicity associated with grating lobes and reduce the number of channels. However, to date, most sparse arrays cannot remove a very large number of channels without negatively affecting imaging quality, most notably by increasing sidelobe levels.

CMUTs are an electrostatic based approach to ultrasound transducers, and are a newer technology than piezoelectrics. By using silicon based fabrication techniques, they can easily be fabricated in many different dense array

architectures, and, in addition, they offer increased bandwidth in immersion operation, and better coupling to external media without the need for impedance matching layers. However, the complexity required to wire every element in a dense 2-D array remains. As such, we have introduced a row-column based architecture that we call top-orthogonal-to-bottom (TOBE) arrays that significantly decreases the number of electronic channels required – to a much greater extent than sparse arrays. By connecting bottom electrodes in columns and top electrodes in orthogonal rows, only 2N channels are required for an N x N array. CMUTs are the superior technology for this array architecture due to their nonlinear operation characteristics. The ability to vibrate with larger amplitudes when subject to a strong DC bias allows the implementation of a number of novel imaging schemes using the row-column architecture that would not be possible with piezoelectric devices.

We have successfully fabricated multiple TOBE arrays using a sacrificial release based fabrication process, and have designed a wafer-bonding process to fabricate them as well. The wafer-bonding process offers significant advantages over the sacrificial release process which could potentially result in higher quality, more uniform arrays. Three different imaging schemes were presented, each offering different trade-offs between imaging speed and imaging resolution for use in varying applications, but each requiring only 2N channels. The flexibility in interface design could allow for further imaging schemes as well. Tests were done to verify the ability of the arrays to perform single-element actuation in both transmitting and receiving modes. Results for both air-coupled and immersion operation showed that in both these modes, element actuation was largely limited to the desired active area, which could be controlled by varying the columns on which the bias voltages were applied. Simulations validated the theory behind the proposed imaging schemes and demonstrated potential for high-resolution imaging using the row-column architecture.

6.2 Future Directions

Future work on the TOBE arrays will center on integrating the arrays with an imaging system and validating their actual imaging capabilities. While individual element testing shows promising results, the effectiveness of the arrays will not be fully demonstrated without demonstrating the ability to obtain actual images. Various factors that can only be tested during actual imaging must be examined. These could include determining whether the signal-to-noise ratio of imaging schemes 1 and 2 can be good enough to obtain high quality images, or if cross-talk non-idealities can propagate through the matrix operations in scheme 3 to result in a worse-than-expected image, among others factors. Presently-fabricated arrays will be packaged and integrated with interface electronics that are currently in development, but should imaging tests provide encouraging results, larger arrays could be fabricated.

In addition, future work will focus on further developing novel imaging schemes using the row-column architecture. Because this architecture is very new and has not been studied for a significant amount of time, especially using CMUTs, the possibilities for introducing never-before-seen schemes are plentiful.

For example, further study could be done on the S-matrix addressing scheme to potentially remove redundancies and increase imaging speed. TOBE arrays could provide new advancements in photoacoustic imaging as well. For example, a novel form of frequency-modulation based addressing could be implemented to potentially allow for the receiving of information from every single individual element in the array using only one laser pulse event. Overall, TOBE CMUTs present an exciting realm of new possibilities for both CMUT technology and ultrasound technology in general, and it is hopeful that the preliminary work done in this thesis will lay the groundwork for many successful improvements in the future.

References

[1] L. Poissonnier et al, "Control of Prostate Cancer by Transrectal HIFU in 227 Patients," *European Urology*, vol. 51, no. 2, pp. 381-387, Feb 2007.

[2] B.T. Khuri-Yakub, O. Oralkan, and M. Kupnik, "Next-gen ultrasound," *IEEE Spectrum*, vol. 46, no. 5, pp. 44-54, May 2009.

[3] Richard Cobbold, *Foundations of Biomedical Ultrasound*. New York, USA: Oxford University Press, 2007.

[4] J.D. Hamilton and M. O'Donnell, "High frequency ultrasound imaging with optical arrays," *IEEE Trans. Ultrason. Ferroelect. Freq. Control*, vol. 45, no. 1, pp. 216-235, Jan 1998.

[5] David Alleyne and Peter Cawley, "The use of Lamb waves for the long range inspection of large structures.," *Ultrasonics*, vol. 34, no. 2-5, pp. 287-290, June 1996.

[6] E.G. Williams and J.D. Maynard, "Numerical evaluation of the Rayleigh integral for planar radiators using the FFT," *J. Acoust. Soc. Am.*, vol. 72, pp. 2020-2030, 1982.

[7] Paul G. Newman and Grace S. Rozycki, "The History of Ultrasound," *Surgical Clinics of North America*, vol. 78, no. 2, pp. 179-195, April 1998.

[8] James F. Tressler, Sedat Alkoy, Robert E. Newnham, "Piezoelectric Sensors and Sensor Materials," *J. Electroceramics*, vol. 2, no. 4, pp. 257-272, Dec 1998.

[9] J K.K. Shung,.M. Cannata and Q.F. Zhou "Piezoelectric materials for high frequency medical imaging applications: A review," *J. Electroceramics*, vol. 19, no. 1, pp. 139-145, Sep 2007.

[10] J.A. Brown et al., "Fabrication and Performance of a 40-MHz Linear Array Based on a 1-3 Composite with Geometric Elevation Focusing," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 54, no. 9, pp. 1888-1894, Sep 2007.

[11] J.M. Reid and J.J. Wild, "Ultrasonic ranging for cancer diagnosis," *Electronics*, vol. 25, no. 5, pp. 136-138, 1952.

[12] O.T. von Ramm, and S.W. Smith, "Beam steering with linear arrays," *Biomed. Eng. IEEE Transactions on,* vol. 30, pp. 438-452, 1983.

[13] G. Baumand and I. Greenwood, "Orbital lesion localization by three dimensional ultrasonography," *NY State J. Med*, vol. 61, pp. 4149-4157, 1961.

[14] A. Fenster and D.B. Downey, "3-D ultrasound imaging: a review."," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 15, no. 6, pp. 41-51, Nov/Dec 1996.

[15] A. Fenster, D.B. Downey and H.N. Cardinal, "Three-dimensional ultrasound imaging," *Phys. Med. Biol*, vol. 46, pp. R67-R99, 2001.

[16] A. Osman et al, "An automated data processing method dedicated to 3D ultrasonic non destructive testing of composite pieces," in *Ultrasound in the Control of Industrial Processes, International Symposium on*, 2012.

[17] D.H. Turnbull and Stuart Foster, "Beam steering with pulsed two-dimensional transducer arrays," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 38, no. 4, pp. 320-333, July 1991.

[18] G.R. Lockwood, J.R. Talman, and S.S. Brunke "Real-Time 3-D Ultrasound Imaging Using Sparse Synthetic Aperture Beamforming," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 45, no. 4, pp. 980-988, July 1998.

[19] S.W. Smith, H.G. Pavy, and O.T. von Ramm "High-Speed Ultrasound Volumetric Imaging System - Part I: Transducer Design and Beam Steering," *IEEE Trans. Ultrason. Ferroelect. Freq. Control*, vol. 38, no. 2, pp. 100-108, March 1991.

[20] O. Oralkan et al, "Volumetric ultrasound imaging using 2-D CMUT arrays," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 50, no. 11, pp. 1581-1984, Nov 2003.

[21] C.E. Demore et al., "Real-Time Volume Imaging Using a Crossed Electrode Array," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 56, no. 6, pp. 1252-1261, June 2009.

[22] D.H. Turnbull and F.S. Foster, "Fabrication and Characterization of Transducer Elements in Two-Dimensional Arrays for Medical Ultrasound Imaging," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 39, no. 4, pp. 464-475, July 1992.

[23] B.T. Khuri-Yakub and O. Oralkan, "Capacitive micromachined ultrasonic transducers for medical imaging and therapy," *J. Micromech. Microeng*, vol. 21, 2011.

[24] S.I Nikolov and J.A. Jensen, "Investigation of the feasibility of 3D Synthetic Aperture Imaging," in *Proc. IEEE Int. Ultrasonics Symp.,* 2003, pp. 1903-1906.

[25] K.E. Thomenius, "Evolution of ultrasound beamformers," in *Proc. IEEE Int. Ultrasonics Symp.,* 1996, pp. 1615-1622.

[26] D.G. Wildes et al, "Elevation performance of 1.25D and 1.5D transducer arrays," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 44, no. 5, pp. 1027-1037, September 1997.

[27] G.R. Lockwood and R.S. Foster, "Design of Sparse Array Imaging Systems," in *Proc. IEEE Int. Ultrasonics Symp.,* 1995, pp. 1237-1243.

[28] Andreas Austeng and Sverre Holm, "Sparse 2D Arrays for Real-Time 3D Ultrasound," in *Proc. Norwegian Signal Processing. Conf.*, Asker, 1999.

[29] G.R. Lockwood and F.S. Foster, "Optimizing the radiation pattern of sparse periodic two-dimensional arrays," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 43, no. 1, pp. 15-19, January 1996.

[30] A. Austeng and S. Holm, "Sparse 2-D Arrays for 3-D Phased Array Imaging - Design Methods," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 49, no. 8, pp. 1073-1086, August 2002.

[31] J.T. Yen, J.P. Steinbert, and S.W. Smith "Sparse 2-D array design for real time rectilinear volumetric imaging," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 47, no. 1, pp. 93-110, Jan 2000.

[32] P.K. Weber et al, "1D- and 2D-Sparse Array Optimization," *Instrumentation Science & Technology*, vol. 27, no. 4, pp. 235-246, 1999.

[33] J.T. Ylitalo and Helmut Ermert, "Ultrasound synthetic aperture imaging: monostatic approach," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 41, no. 3, pp. 333-339, May 1994.

[34] S.Bennett et al, "A real-time synthetic aperture digital acoustic imaging system," *Acoust. Imaging*, vol. 10, pp. 669-692, 1982.

[35] S.I. Nikolov, "Synthetic aperture tissue and flow ultrasound imaging," Center for Fast Ultrasound Imaging, Technical University of Denmark, Lyngby, Ph D Dissertation 2001.

[36] S.I. Nikolov and J.A., Jensen, "3D synthetic aperture imaging using a virtual source element in the elevation plane," in *Proc. IEEE Int. Ultrasonics Symp.,* 2000, pp. 1743-1747.

[37] N.M. Daher and J.T. Yen, "2-D Array for 3-D Ultrasound Imaging Using Synthetic Aperture Techniques," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 53, no. 5, pp. 912-924, May 2006.

[38] O.Oralkan et al, "Capacitive micromachined ultrasonic transducers: next-generation arrays for acoustic imaging," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 49, no. 11, pp. 1596-1610, Nov 2002.

[39] Peiyu Zhang et al, "Double-SOI wafer-bonded CMUTs with improved electrical safety and minimal roughness of dielectric and electrode surfaces," *Microelectromechanical Systems*, vol. 21, no. 3, pp. 668-680, June 2012.

[40] A. Caronti et al, "Capacitive micromachined ultrasonic transducer (CMUT) arrays for medical imaging," *Microelectronics Journal*, vol. 37, no. 8, pp. 770-777, August 2006.

[41] Kumar Tiwari et al, "Hexagonal capacitance micromachined ultrasonic transducer," in *Computing, Communication and Networking, ICCCn Internation Conference on,* 2008, pp. 1-4.

[42] B.T. Khuri-Yakub et al, "Silicon micromacined ultrasound transducers," *Jpn. J. Appl. Phys.*,

vol. 39, pp. 2883-2887, May 2000.

[43] Stephen D. Senturia, *Microsystem Design.*: Springer, 2001.

[44] I.O. Wygant, M. Kupnik, and B.T. Khuri-Yakub, "Analytically calculating membrane displacement and the equivalent circuit model of a circular CMUT cell," in *I Proc. IEEE Int. Ultrasonics Symp.,* 2008, pp. 2111-2114.

[45] G.G. Yaralioglu et al., "Calculation and measurement of electromechanical coupling coefficient of capacitive micromachined ultrasonic transducers," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 50, no. 4, pp. 449-456, April 2003.

[46] W. Zheng et al, "Empirical model for dielectric charging in double-SOI-wafer-bonded CMUTs: Theory and experiment," in *Proc. IEEE Int. Ultrasonics Symp.,* 2011, pp. 353-356.

[47] A. Kshirsagar, A.Sampaleanu, R. Chee, W. Moussa, and R. Zemp, "Pre-charged CMUTs with efficient low-bias voltage operation for medical applications," in *Proc. IEEE Int. Ultrasonics Symp.,* 2013, p. Submitted.

[48] T.J. Reck, J.A. Hossack, M.D.C Eames, "Selectable frequency CMUT with membrane stand-off structures," in *Proc. IEEE Int. Ultrasonics Symp.,* 2009.

[49] X.C. Jin et al, "Characterization of one-dimensional capacitive micromachined ultrasonic immersion transducer arrays," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 48, no. 3, pp. 750-760, May 2001.

[50] S. Vaithilingam et al, "Three-dimensional photoacoustic imaging using a two-dimensional CMUT array," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 56, no. 11, pp. 2411-2419, Nov 2009.

[51] A.S. Ergun et al, "Capacitive micromachined ultrasonic transducers: fabrication technology," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 52, no. 12, pp. 2242-2258, Dec 2005.

[52] Y. Huang et al, "A solution to the charging problems in capacitive micromachined ultrasonic transducers," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 52, no. 4, pp. 578-580, April 2005.

[53] M.I Haller and B.T Khuri-Yakub, "A surface micromachined electrostatic ultrasonic air transducer," in *Proc. IEEE Int. Ultrasonics Symp.,* 1994, pp. 1241-1244.

[54] D. Spoliansky, I. Ladabaum, and B.T. Khuri-Yakub "Micromachined ultrasonic air transducers (MUTs)," *Microelectric Engineering*, vol. 30, pp. 535-538, Jan 1996.

[55] M. Yew et al, "A study of residual stress effects on CMOS-MEMS microphone technology," in *Impact Conf.*

[56] R.A. Noble et al, "Low temperature micromachined CMUTs with fully-integrated analogue front-end electronics," in *Proc. IEEE Int. Ultrasonics Symp.,* 2002, pp. 1045-1050.

[57] E.P. van de Ven, I.W. Connick, and A.S. Harrus "Advantages of dual frequency PECVD for deposition of ILD and passivation films," in *VLSI Multilevel Interconnection Conference, IEEE*, 1990, pp. 194-201.

[58] J. Knight et al, "Low temperature fabrication of immersion capacitive micromachined ultrasonic transducers on silicon and dielectric substrates," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 51, pp. 1324-1333.

[59] N.I. Maluf, K.E. Petersen, G.T. Kovacs, "Bulk micromachining of silicon," *Proceedings of the IEEE*, vol. 86, no. 8, pp. 1536-1551, Aug 1998.

[60] K.R. Williams and R.S. Muller, "Etch rates for micromachining processing," *J. MEMS*, vol. 5, no. 4, pp. 256-269, Dec 1996.

[61] Younsoo Park et al., "Evolution of residual stress in plasma-enhanced chemical-vapor-deposited silicon dioxide film exposed to room air," *Appl. Phys. Lett.*, vol. 75, no. 24, pp. 3811-3813, 1999.

[62] P.B Chu et al, "Controlled pulse-etching with xenon difluoride," in *Solid State Sensors and Actuators,* Chicago, 1997, pp. 665-668.

[63] Y. Huang et al, "Fabricating capacitive micromachined ultrasonic transducers with wafer-

bonding," *J.MEMS*, vol. 12, pp. 128-137, 2003 2003.

[64] D.S Lin et al, "Characterization of fabrication related gap-height variations in capacitive micromachined ultrasonic transducers," in *Proc. IEEE Int. Ultrasonics Symp.,* 2007, pp. 523-526.

[65] M. Kupnik et al, "CMUT fabrication based on a thick buried oxide layer," in *Proc. IEEE Int. Ultrasonics Symp.,* 2010, p. Submitted.

[66] Y.Huang et al, "Capacitive micromachined ultrasound transducers with piston-shaped membranes: fabrication and experimental characterization," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,*

[67] P.C Eccardt et al, "Surface micromachined ultrasound transducers in CMOS technology," in *Proc. IEEE Int. Ultrasonics Symp.,* 1996, pp. 959-62.

[68] Y.Tsuji, M. Kupnik, and B.T. Khuri Yakub "Low temperature process for CMUT fabrication with wafer bonding technique," in *Proc. IEEE Int. Ultrasonics Symp.,* 2010.

[69] D.S Lin et al, "Encapsulation of capacitive micromachined ultrasonic transducers using viscoelastic polymer," *J.MEMS*, vol. 19, pp. 1341-1351, 2010.

[70] A. Savoia et al, "Crosscross 2D cMUT Array: Beamforming Strategy and Synthetic 3D Imaging Results," *Proc. IEEE Int. Ultrasonics Symp.,* pp. 1514-1517, Oct 2007.

[71] M. Pappalardo et al, "Bidimensional Ultrasonic Array for Volumetric Imaging," *U.S. Patent 2010137718 (A1)*, June 17 2008.

[72] Chi Hyung Seo and J.T. Yen, "A 256x256 2-D Array Transducer with Row-Column Addressing for 3-D Imaging," *Proc. IEEE Int. Ultrasonics Symp.,* pp. 2381-2384, Oct 2007.

[73] J.T. Yen et al, "A dual-layer transducer array for 3-D rectilinear imaging," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 56, no. 1, pp. 204-212, Jan 2009.

[74] A.S. Logan et al, "A 32 x 32 element row-column addressed capacitive micromachined ultrasonic transducer," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 58, no. 6, pp. 1266-1271, June 2011.

[75] A.S. Logan, L.L. Wong, and J.T.W. Yeow, "2-D CMUT wafer bonded imaging arrays with a row-column addressing scheme," *Proc. IEEE Int. Ultrasonics Symp.,* pp. 984-987, Sept 2009.

[76] Peiyu Zhang et al, "Double-SOI Wafer-Bonded CMUTs With Improved Electrical Safety and Minimal Roughness of Dielectric and Electrode Surfaces," *Microelectromechanical Systems, Journal of*, vol. 21, no. 3, pp. 668-680, June 2012.

[77] B. Bayram et al, "Capacitive Micromachined Ultrasonic Transducer Design for High Power Transmission," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 52, no. 2, pp. 326-339, Feb 2005.

[78] J.A. Jensen, "Field: A Program for Simulating Ultrasound Systems," *Paper presented at the 10th Nordic-Baltic Conference on Biomedical Imaging, Published in Medical & Biological Engineering & Computing*, vol. 34, no. Supplement 1, Part 1, pp. 351-353, 1996.

[79] J.A. Jensen and N.B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 39, pp. 262-267, 1992.

[80] J.A. Jensen and S.I. Nikolov, "Fast simulation of ultrasound images," in *Proc. IEEE Int. Ultrasonics Symp.,* pp. 1721-1724.

[81] T. Harrison and R.J. Zemp, "The applicability of ultrasound dynamic receive beamformers to photoacoustic imaging," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 58, no. 11, pp. 2259-2263, Oct 2011.

[82] W. Lee et al, "A miniaturized catheter 2-D array for real-time, 3-D intracardiac echocardiography," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 51, no. 10, pp. 1334-1346, Oct 2004.

[83] D. Dausch et al, "Theory and operation of 2-D array piezoelectric micromachined ultrasound transducers," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 55, no. 11, pp. 2484-

2492, Nov 2008.

[84] B. Bayram et al, "Finite element modeling and experimental characterization of crosstalk in 1-D CMUT arrays," *IEEE Trans. Ultrason. Ferroelect. Freq. Control,* vol. 54, no. 2, pp. 418-430, Feb 2007.

Appendix 1

Synthetic Aperture 3D Ultrasound Imaging Schemes witih S-Sequence Bias-Encoded Top-Orthogonal-to-Bottom-Electrode 2D CMUT Arrays.

Alexander Sampaleanu, Roger Zemp*
Department of Electrical and Computer Engineering
University of Alberta, Edmonton, Canada
*rzemp@ualberta.ca

***Abstract*—Crossed electrode arrays, wherein electrodes are connected in individually addressable rows and columns, have been previously investigated as a means to reduce the number of channels in two-dimensional ultrasound arrays for three-dimensional imaging. Previously envisioned imaging schemes have been limited in their resolution and focusing ability due to their inability to perform active focusing in both the lateral and elevational direction for both transmit and receive. As well, signal-to-noise ratio (SNR) can be compromised when transmitting or receiving on only one row or column at a time. A new imaging method is proposed for this array architecture using capacitive micromachined ultrasound transducers (CMUTs) which can provide two-way focusing in both lateral and elevational directions while providing optimal SNR. This method is a variation of synthetic aperture imaging wherein the columns active upon transmit and receive events are determined by applying a bias to specific columns upon transmission and reception in an S-sequence encoded pattern. Mathematical derivations of this method are presented and simulations are done to compare this imaging method to previously published crossed-electrode array imaging schemes, as well as to fully-populated 2-D phased array imaging.**

Keywords—3-D imaging; 2-D array; CMUTs; Crossed-Electrode Array;

I.Introduction

While there is a strong desire for high quality 3-D ultrasound imaging for use in different applications, 2-D array architectures have struggled to develop to the extent necessary to see widespread use. Large apertures display increased resolution, but require many small, densely packed elements to prevent grating lobes at large steering angles. A square 2-D array with N elements per side has N^2 elements, and it becomes expensive and complicated to wire each separate element. The crossed-electrode architecture has been previously investigated as a means to solve this problem. In this architecture, the N x N 2-D array contains N column electrodes and N row electrodes which are orthogonal to each other, as

shown in Fig. 1. We refer to these arrays as 'top-orthogonal-to-bottom-electrode' (TOBE) arrays. In this form, the array structure has been investigated previously by Yen et al with piezoelectric technology [1] [2], and by Logan et al using CMUTs [3] [4]. We have recently expanded on this concept using our own

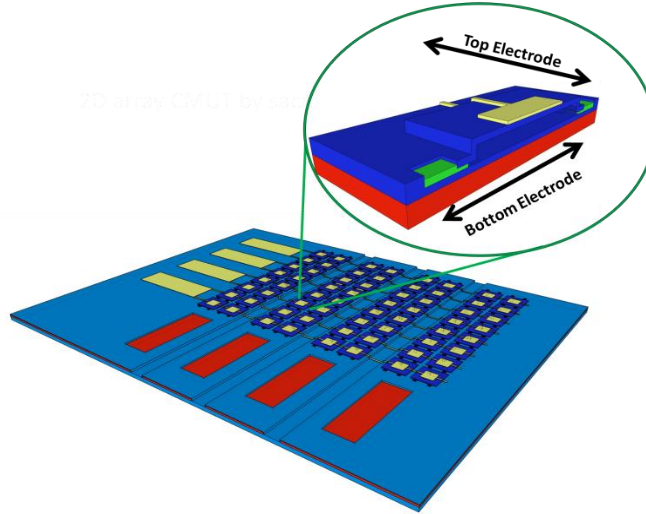


Fig. 1. 3-D visualization of our fabricated TOBE array

TOBE CMUT arrays and introduced new ways of imaging using these arrays [5]. By biasing only one column at a time and routing transmit-receive signals along rows, dominant single-element operation was demonstrated due to the fact that only the element at the intersection of the active row and column experienced both signal and bias. Through acoustic and laser Doppler vibrometer tests, we found that actuation of other elements was negligible. The initially proposed pulse sequence using this addressing scheme was what could be called an 'elevationally-focused classical synthetic aperture' technique. This scheme was similar to the monostatic classical synthetic aperture focusing techniques described in [90] [91], and others, where one element at a time is used for transmitting and receiving. For the TOBE CMUTs, one column at a time was biased to serve as an effective thin linear array which could form elevationally focused but laterally unfocused B-scans. Synthetic aperture methods were then used to retroactively focus the set of B-scans along the lateral direction. The principal disadvantage of this imaging method is that only a single column is active at a time, and there is no active transmit focusing in the lateral direction resulting in a low signal-to-noise ratio. An additional problem lies in the fact that since only one element per row is active at a time, parasitic capacitance from interconnects could swamp the small capacitance of the elements without careful engineering. This could result in additional signal-to-noise degradation, compounding the problem. To address this difficulty, we propose a novel coded aperture pulse sequence that has the potential to significantly improve signal-to-noise ratio and resolution at the cost of additional transmit events.

Previously proposed approaches used the same column bias for both transmit and receive operations. The proposed method involves changing the bias pattern after a transmit event and prior to reception. This must be done in a way that generates negligible transmit signal during bias-switching and must be done during the dead-time after a transmit event. There is typically a dead-time on the order of a microsecond in most systems due to recovery of amplifiers after saturation due to a transmit event. A dead-time of a microsecond or two would lead to ample time to switch bias states of CMUTs. If acoustic transients are generated they would be primarily at low <1MHz frequencies that could be filtered upon reception. Alternatively, a more complicated architecture could be used involving separate rows of devices for transmit and receive.

II.Theory

The pulse-echo signal due to a transmitting element at location $\boldsymbol{r}_k$ and receiver element at location $\boldsymbol{r}_j$ and a scatterer at location $\mathbf{r}$, excluding measurement noise, is given as:

$$e_{kj}(t) = \frac{d}{dt}\rho V_{Tx}(t) * T_{Tx}(t) * T_{Rx}(t) * \ldots$$

$$\ldots \int S(\boldsymbol{r} \;\; ,t) * h_{Tx}(\boldsymbol{r}_k \to \boldsymbol{r} \;\; ,t) * h_{Tx}(\boldsymbol{r}_j \to \boldsymbol{r} \;\; ,t) d\boldsymbol{r}$$

(1)

Now consider that elevationally-focused B/A-scans are obtained using multiple columns out of the TOBE array. The active columns are defined by an on-off binary pattern chosen from binary masks. Let the i_{th} binary mask for a transmit event denoted as $\boldsymbol{w}_{Tx}^{i} = \left[w_{Tx,1}^{i}, w_{Tx,2}^{i}, \ldots, w_{Tx,N}^{i}\right]^{T}$ where $w_{Tx,j}^{i}$ is the binary 0/1 weight applied to the jth TOBE CMUT column and represents whether a bias voltage is present or absent. Likewise, let $\boldsymbol{w}_{Rx}^{i} = \left[w_{Tx,1}^{i}, w_{Tx,2}^{i}, \ldots, w_{Tx,N}^{i}\right]^{T}$ be the weights associated with a receive-acquisition. Then the beamformed pulse-echo signal due to the transmit mask $\boldsymbol{w}_{Tx}^{m}$ and the receive mask $\boldsymbol{w}_{Rx}^{p}$ is given as:

$$v_{mp}(t) = [\boldsymbol{w}_{Tx}^{m}]^{T} \boldsymbol{E}(t) \boldsymbol{w}_{Tx}^{p} + n_{mp}(t) \quad (2)$$

where $n_{mp}(t)$ is a statistically-independent noise process and $[\boldsymbol{E}]_{kj} = e_{kj}$. The collection of signals due to M transmit masks and P receive masks may be organized into matrix form as

$$\boldsymbol{V}(t) = \boldsymbol{W}_{\boldsymbol{Tx}}^{\boldsymbol{T}} \boldsymbol{E}(t) \boldsymbol{W}_{\boldsymbol{Rx}} + \boldsymbol{n}(t) \quad (3)$$

where

$$\boldsymbol{W_{Rx}} = [\,\boldsymbol{w}_{Rx}^{1}; \boldsymbol{w}_{Rx}^{2}; \dots; \boldsymbol{w}_{Rx}^{M}] \tag{4a}$$

$$\boldsymbol{W_{Tx}} = [\,\boldsymbol{w}_{Tx}^{1}; \boldsymbol{w}_{Tx}^{2}; \dots; \boldsymbol{w}_{Tx}^{M}] \tag{4b}$$

and $\boldsymbol{n}(t)$ is a matrix of statistically independent zero-mean stochastic noise processes with variance σ^2. The goal in this imaging problem is to estimate the pulse-echo signals due to each transmitter-receiver element pair, $\boldsymbol{E}(t)$ based on choice of $\boldsymbol{W_{Tx}}$ and $\boldsymbol{W_{Rx}}$ so as to maximize the signal-to-noise ratio of the estimate $\widehat{\boldsymbol{E}}$ and to minimize the mean-square error**.** Then we can apply full synthetic aperture focusing to obtain the best-possible resolution and best-possible imaging signal-to-noise. In doing so we seek a linear estimator of the form $\widehat{\boldsymbol{E}} = \boldsymbol{AVB}$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices selected so that $\widehat{\boldsymbol{E}}$ is an unbiased estimator that minimizes each element of the error matrix $\boldsymbol{\epsilon} = \langle(\widehat{\boldsymbol{E}} - \boldsymbol{E})(\widehat{\boldsymbol{E}} - \boldsymbol{E})^{\boldsymbol{T}}\rangle$**.** Scalar metrics for minimization could include $\varepsilon = \|\boldsymbol{\epsilon}\|$ or $Tr\,\boldsymbol{\epsilon}$**.**

$\langle\widehat{\boldsymbol{E}}\rangle = \boldsymbol{A}\langle\boldsymbol{V}\rangle\boldsymbol{B} = \boldsymbol{A}\boldsymbol{W}_{\boldsymbol{Tx}}^{\boldsymbol{T}}\langle\boldsymbol{E}\rangle\boldsymbol{W_{Rx}}\boldsymbol{B} + \boldsymbol{A}\langle\boldsymbol{n}\rangle\boldsymbol{B} = \boldsymbol{A}\boldsymbol{W}_{\boldsymbol{Tx}}^{\boldsymbol{T}}\boldsymbol{E}\boldsymbol{W_{Rx}}\boldsymbol{B}$ satisfies unbiasedness $\langle\widehat{\boldsymbol{E}}\rangle = \boldsymbol{E}$ if $\boldsymbol{A}$ and $\boldsymbol{B}$ are be chosen as the Moore-Penrose pseudo-inverses of $\boldsymbol{W}_{\boldsymbol{Tx}}^{\boldsymbol{T}}$ and $\boldsymbol{W_{Rx}}$**,** respectively:

$$\boldsymbol{A} = \left[\boldsymbol{W_{Tx}}\boldsymbol{W}_{\boldsymbol{Tx}}^{\boldsymbol{T}}\right]^{-\boldsymbol{1}}\boldsymbol{W_{Tx}} \tag{5a}$$

$$\boldsymbol{B} = \boldsymbol{W}_{\boldsymbol{Rx}}^{\boldsymbol{T}}\left[\boldsymbol{W_{Rx}}\boldsymbol{W}_{\boldsymbol{Rx}}^{\boldsymbol{T}}\right]^{-\boldsymbol{1}} \tag{5b}$$

In this case $\widehat{\boldsymbol{E}} - \boldsymbol{E} = \boldsymbol{AVB} - \boldsymbol{E} = \boldsymbol{AnB}$ so

$$\boldsymbol{\epsilon} = \langle\boldsymbol{AnB}(\boldsymbol{AnB})^{\boldsymbol{T}}\rangle = \boldsymbol{A}\langle\boldsymbol{nBB^{T}n^{T}}\rangle\boldsymbol{A^{T}} \tag{6}$$

III.Determining the weighting matrices

At this point we consider several options for weighting matrices $\boldsymbol{W_{Tx}}$ and $\boldsymbol{W_{Rx}}$**.** If both $\boldsymbol{W_{Tx}}$ and $\boldsymbol{W_{Rx}}$ are selected as NxN identity matrices $\boldsymbol{I_N}$, where N is the size of the array, then this corresponds to classical synthetic aperture acquisition sequences where one element at a time fires and another element is used to receive signal. In this case $\boldsymbol{\epsilon} = \langle\boldsymbol{nn^{T}}\rangle = N\sigma^2\boldsymbol{I}$, so $Tr\boldsymbol{\epsilon}/\sigma^2 = N^2$ and $\frac{\|\boldsymbol{\epsilon}\|}{\sigma^2} = N^2$.

Next, consider taking both $\boldsymbol{W_{Tx}}$ and $\boldsymbol{W_{Rx}}$ to be NxN Hadamard S-matrices, $\boldsymbol{S}$. The S-Matrix is formed by taking an MxM Hadamard matrix of order $M=2^n$ $=N+1$ in which the entries in the first row and column are 1, changing ones to zeros and -1's to ones, and deleting the first row and column. The rows or columns of the S-Matrix are sometimes called S-sequences and would represent the biasing scheme applied for each transmit and/or receive event. It has been shown [8] that $\boldsymbol{SS^{T}} = \frac{N+1}{4}(\boldsymbol{I} + \boldsymbol{J})$, where $\boldsymbol{J}$ is an NxN matrix of ones. Likewise,

$\boldsymbol{S^{-1}} = \frac{2}{N+1}[2\boldsymbol{S^T} - \boldsymbol{J}]$ and $\boldsymbol{SJ} = \boldsymbol{S^T J} = \frac{N+1}{2}\boldsymbol{J}$. With these results, one can show that the matrix condition number κ, gaging the stability of the matrix inversion and assuming L2-norms, is then given as $\kappa(\boldsymbol{S_N}) = \sqrt{N+1}$ which is of course worse than the $\kappa(\boldsymbol{I}) = 1$ for the column-by-column addressing scheme but much more asymptotically attractive than a good many other choices for $\boldsymbol{W}$. We may also compute

$$\boldsymbol{S}^{-1}\boldsymbol{S}^{-\boldsymbol{T}} = \left(\frac{2}{N+1}\right)^2 [(N+1)\boldsymbol{I} - \boldsymbol{J}] \tag{7}$$

We can then combine equation (6) with equations (5a) and (5b) expressed in terms of equation (7) to obtain:

$$\frac{\epsilon}{\sigma^2} = N^2\left(\frac{2}{N+1}\right)^4 [(N+1)\boldsymbol{I} - \boldsymbol{J}] \tag{8}$$

Diagonal elements of this error matrix have magnitude of order $\mathcal{O}\{N^3/(N+1)^4\} \sim \mathcal{O}\{1/N\}$ which is a factor of N^2 better than the $\mathcal{O}\{N\}$ associated with identity weighting matrices (traditional synthetic aperture scanning) while off-diagonal elements are of order $\mathcal{O}\{N^2/(N+1)^4\} \sim \mathcal{O}\{1/N^2\}$ which can be very small for large N but slightly worse than the value of zero associated with traditional synthetic aperture scanning. Overall trace and L2-norm scalar error metrics are lower for the new scheme. For example,

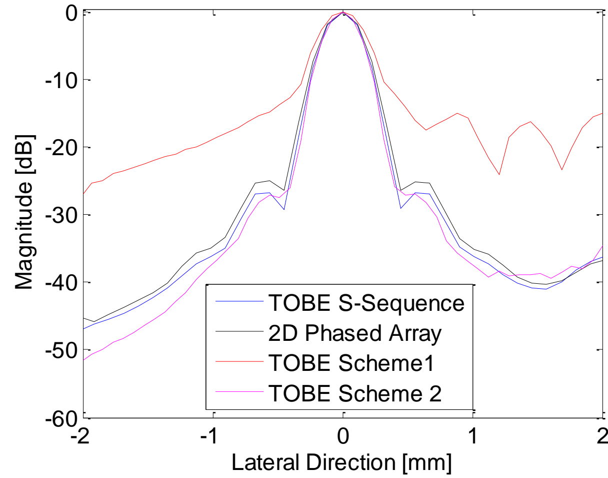
$\frac{Tr\epsilon}{\sigma^2} = N^4\left(\frac{2}{N+1}\right)^4$ compared to N^2 for traditional synthetic aperture scanning, a significant improvement for large N. Additionally $\frac{\|\epsilon\|}{\sigma^2} = N^2\left(\frac{2}{N+1}\right)^4 (N+1)$ for the S-Matrix aperture encoding compared with N^2 for traditional synthetic aperture scanning. For a 127 x 127 array, S-matrix aperture encoding improves $\frac{Tr\epsilon}{\sigma^2}$ by a factor of more than 1000, and improves $\frac{\|\epsilon\|}{\sigma^2}$ by a factor of more than 100,000.

Because for each transmit/receive S-sequence, there are (N+1)/2 columns biased at a time, this scheme offers up to (N+1)/2 times the amount of transmit signal and up to (N+1)/2 times the amount of receive signal over N bias sequences compared to our previously suggested TOBE imaging schemes. In the matrix inversion operation, noise propagation reduces the actual SNR gain in individual elements.

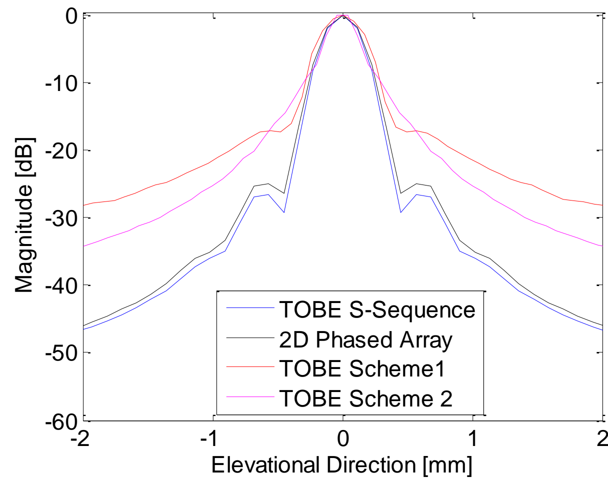
Recovery of the data, $\widehat{\boldsymbol{E}} = \boldsymbol{S^{-1}VS^{-T}}$, is facilitated if the number of elements encoded is $N = 2^n - 1$. In these cases, the fast Hadamard transform (FHT) can be used instead of direct application of the inverse matrix, reducing the number of steps from $n^2 - n$ to $n\log_2 n$ for each matrix multiplication. Because this must be done for each time-point, such computational savings could be particularly important for real-time imaging.

IV.Simulations

Simulations have been done using Field II [9] with the aid of the beamforming toolbox developed by Jensen and Nikolov [94]. A 63 x 63 element TOBE CMUT was used for the S-sequence imaging simulations, and this was compared to a fully-wired 2D array of the same size, as well as to the previous two TOBE array imaging schemes modeled in [5].



a)



b)

Fig. 2: Cross-Range-Maximum-Amplitude point-spread function plots comparing the TOBE S-sequence imaging method with a fully-wired 2D phased array as well as two previously proposed TOBE imaging methods in a)The lateral direction and b) the elevational direction

The maximum amplitude projection C-scan image at an imaging depth of 104 wavelengths from the arrays surface was found for all modeled imaging schemes, after which the cross-section of these c-scans in both lateral and elevational directions were plotted. These cross sections are shown in Fig.2 Afterwards, the

same simulations were repeated, but a constant noise level was added to all the simulations so as to compare the SNR of each scheme. In addition, a simulation was done where the weighting matrices $\boldsymbol{W_{Tx}}$ and $\boldsymbol{W_{Rx}}$ were identity matrices. The relative SNRs are given in Table 1.

V.Discussion

Based on the simulation results, it is clear that the S-matrix aperture encoded imaging scheme we propose shows considerable potential in terms of imaging quality. The point spread function plots show that the resolution is comparable to – if not slightly better than – a fully-wired 2D phased array of the same size. At the same time, because of the crossed-electrode design which requires much fewer channels than a fully-wired array, such an imaging scheme would be much less complicated to implement. When compared to previous imaging schemes proposed for the crossed-electrode architecture, both by our group and others, the imaging scheme shows considerable improvement in both lateral and elevational resolution. In addition, the signal-to-noise ratio is vastly improved over previously proposed crossed-electrode imaging schemes. When compared to an aperture encoding scheme using identity weighting matrices, the SNR is increased by approximately 12 times, which is on the same order of magnitude as predicted by equation 9.

Significant points of consideration regarding this imaging scheme involve the bias voltage switching time and the imaging speed. As previously mentioned, because this imaging scheme involves bias voltage switching between transmit and receive events, the switching time would need to be fast enough to accomplish this, but also slow enough to mitigate unwanted transmit signals during bias switching. This would require careful engineering of the accompanying electronics. There are commercially available high-voltage switches with ~µs switching times which could potentially achieve fast enough switching rates. Additionally, previously published literature has shown that CMUT membrane response to applied bias voltage can be fast enough to keep up with fast switching times [11]. A drawback of this proposed imaging scheme in its current form would be the frame rate. The number of required transmit events for an N x N array would be N x N x S_Y, which is N times slower than our previously proposed TOBE imaging scheme (Scheme 2). It would also be slower than any fully-wired 2D array that did not use an unnecessarily large number of scanlines. This likely means that the S-matrix encoded imaging scheme would not be suitable for real-time medical imaging applications.

SNR COMPARISONS BETWEEN VARIOUS 3-D IMAGING SCHEMES

Imaging Scheme	**SNR**
TOBE S-Matrix Imaging Scheme	+48.3 dB
2D Phased Array	+40.3 dB

Identity Weighting Matrix Scheme	+26.5 dB
Previously Proposed TOBE Scheme 1	+11.1 dB
Previously Proposed TOBE Scheme 2	+9.9 dB

However, the ability to provide high quality images with a less complicated design than fully populated 2D arrays could make it useful for industrial applications such as NDT, where imaging speed is not as critical.

VI.Conclusion and Future Work

The ability to provide imaging quality equal to fully-wired 2-D phased arrays, which have been considered as the 'gold-standard' of 3-D imaging, without the complex fabrication requirements, has been a topic of research for some time now. Simulations of our proposed imaging scheme show its potential to provide high-resolution 3D images with optimal signal-to-noise ratio while requiring only 2N channels for an N x N array. Future work will involve applying this imaging scheme to our fabricated arrays and performing acoustic tests to validate the simulations. In addition, further study will be performed on optimal encoding schemes to reduce redundancies and improve imaging speeds.

References

[1] Chi Hyung Seo and J.T. Yen, "A 256x256 2-D Array Transducer with Row-Column Addressing for 3-D Imaging," in *Ultrasonics Symposium, 2007. IEEE*, 2007, pp. 2381-2384.

[2] J.T. Yen, C.H. Seo, S.I. Awad, J.S. Jeong, "A dual-layer transducer array for 3-D rectilinear imaging," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, vol. 56, no. 1, pp. 204-212, Jan 2009.

[3] A.S. Logan, L.P. Wong, A.Chen, J.T. Yeow "A 32x32 element row-column addressed capacitive micromachined ultrasonic transducer," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, vol. 58, no. 6, pp. 1266-1271, June 2011.

[4] A.S. Logan, L.L. Wong, and J.T. Yeow "2-D CMUT wafer bonded imaging arrays with a row-column addressing scheme," in *International Ultrasonics Symposium, IEEE*, 2009, pp. 984-987.

[5] A.Sampaleanu, P. Zhang, A. Kshirsagar, W. Moussa, R.J. Zemp, "Top-orthogonal-to-bottom-electrode arrays (TOBE) CMUT arrays for 3-D imaging," *Submitted to IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*.

[6] S.Bennett et al, "A real-time synthetic aperture digital acoustic imaging system," *Acoust. Imaging*, vol. 10, pp. 669-692, 1982.

[7] J.T. Ylitalo and Helmut Ermert, "Ultrasound synthetic aperture imaging: monostatic approach," *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*, vol. 41, no. 3, pp. 333-339, May 1994.

[8] T.Fine, P.G. Phillips, and M.Harwitt N.Sloane, "Codes for multiplex spectrometry," *Applied Optics*, vol. 8, no. 10, Oct 1969.

[9] J.A. Jensen, "Field: A Program for Simulating Ultrasound Systems," *Medical & Biological Engineering & Computing*, vol. 34, no. Supplement 1, Part 1, pp. 351-353, 1996.

[10] J.A. Jensen and S.I. Nikolov, "Fast Simulation of Ultrasound Images," in *Proc. IEEE Ultrason. Symp.*, 2000, pp. 1721-1724.

[11] B. Bayram et al, "Capacitive Micromachined Ultrasounic Transducer Design for High Power Transmission," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on,* vol. 52, no. 2, pp. 326-339, Feb 2005.