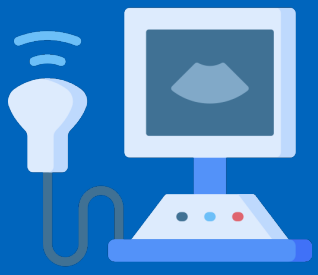
Computer Aided Medical Procedures
Prof. Dr. Nassir Navab



Dissertation

# Physics-Informed Deep Learning for Advanced Medical Ultrasound

Walter Arthur Simson IV

TUM School of Computation,
Information and Technology
Technische Universität München

Technische Universität München
TUM School of Computation, Information and Technology

# Physics-Informed Deep Learning for Advanced Medical Ultrasound

Walter Arthur Simson

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

*Vorsitzende(r):* Prof. Dr. Julien Gagneur

*Prüfer der Dissertation:* 1. Prof. Dr. Nassir Navab

2. Prof. Dr. Robert Rohling

Die Dissertation wurde am 14.03.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 04.11.2022 angenommen.

**Walter Arthur Simson IV**
*Physics-Informed Deep Learning for Advanced Medical Ultrasound*
Dissertation, Version 2.0

**Technische Universität München**
TUM School of Computation, Information and Technology
Lehrstuhl für Informatikanwendungen in der Medizin
Boltzmannstraße 3
85748 and Garching bei München

# Abstract

Medical ultrasound imaging is safe, portable, and inexpensive and can aid clinicians in diagnosing and identifying many early disease symptoms. Ultrasound imaging is based on the pulse-echo principle of transmission and reception of acoustic ultrasound waves in human tissue without the use of harmful ionizing radiation. Despite the many advantages of ultrasound imaging, ultrasound images can suffer from low signal-to-noise ratio and low contrast, potentially hindering clinical interpretation. Some of these image quality challenges result from assumptions about quantitative value distributions in the interrogated medium, such as sound speed. The assumption of a constant sound speed of 1540 m/s in a heterogeneous medium can lead to a loss of resolution and added noise in the resulting ultrasound B-modes. Recently, new algorithms have been developed in quantitative ultrasound, which improve ultrasound image quality by estimating the quantitative composition of the medium. To this end, this dissertation will present and discuss a new method of sound speed estimation with deep neural networks. Accurate ultrasound simulations are used to train a deep neural network to learn a mapping from complex in-phase and quadrature component ultrasound signals to a spatial distribution of sound speed. The network is trained on the results of accurate and realistic ultrasound simulations and evaluated on real-world phantoms and in-vivo study data.

This work is structured in three parts, an introduction of physical and deep learning principles, a discussion of ultrasound simulations and their parametrization, and a presentation of sound speed estimation with deep learning. In Part 1, we will discuss the physical principles of ultrasound to understand how ultrasonic waves are generated, propagated, and are received in medical imaging. We will cover the basics of the wave equation, the concepts of attenuation, absorption, non-linearity, reflection, and scatterer statistics. We will subsequently discuss the fundamental principles of deep learning and their application in ultrasound imaging. In Part 2, we will discuss methods to parameterize and generate accurate and robust in silico ultrasound simulations. The ability to represent the natural processes of ultrasound wave propagation and sub-wavelength scattering in computational models is critical for a physics-informed neural network system. We show that a realistic in-silico phantom can be prepared as the input for a numerical ultrasound simulation with tissue property values from the literature. The proposed simulation method in this work can achieve a high level of realism and enable sound speed estimation. In Part 3, we present a new and novel method for sound speed estimation in clinical breast ultrasound images. This method builds on the physical and deep learning fundamentals and the simulation techniques of the previous two Parts. We show the effective use of deep neural networks to estimate sound speed maps in phantom and real-world volunteer data. This work could significantly impact clinical outcomes by improving downstream image quality and adding a quantitative metric of sound speed on which diagnostic models can be built.

# Zusammenfassung

Die medizinische Ultraschallbildgebung ist sicher, handlich und kostengünstig und kann Ärzten bei der Diagnose und Erkennung vieler früher Krankheitssymptome helfen. Die Ultraschallbildgebung basiert auf dem Impuls-Echo-Prinzip der Übertragung und des Empfangs akustischer Ultraschallwellen in menschlichem Gewebe, ohne dass dabei schädliche ionisierende Strahlung zum Einsatz kommt. Trotz der vielen Vorteile der Ultraschallbildgebung können die Ultraschallbilder ein geringes Signal-Rausch-Verhältnis und einen geringen Kontrast aufweisen, was die klinische Interpretation möglicherweise erschwert. Einige dieser Probleme mit der Bildqualität sind das Ergebnis von Annahmen über quantitative Werteverteilungen im untersuchten Medium, wie z. B. der Schallgeschwindigkeit. Die Annahme einer konstanten Schallgeschwindigkeit von 1540 m/s in einem heterogenen Medium kann zu einem Auflösungsverlust und zusätzlichem Rauschen in den resultierenden Ultraschall-B-Modes führen. In jüngster Zeit werden auf dem Gebiet des quantitativen Ultraschalls neue Algorithmen entwickelt, die die Qualität der Ultraschallbilder verbessern, indem sie die quantitative Zusammensetzung des Mediums abschätzen. Zu diesem Zweck wird in dieser Dissertation eine neue Methode zur Schätzung der Schallgeschwindigkeit mit tiefen neuronalen Netzen vorgestellt und diskutiert. Anhand genauer Ultraschallsimulationen wird ein tiefes neuronales Netz trainiert, um eine Abbildung komplexer In-Phase- und Quadraturkomponenten von Ultraschallsignalen auf eine räumliche Verteilung der Schallgeschwindigkeit zu lernen.

Diese Arbeit ist in drei Teile gegliedert: eine Einführung in die physikalischen und Deep-Learning-Prinzipien, eine Diskussion der Ultraschallsimulationen und ihrer Parametrisierung sowie eine Darstellung der Schallgeschwindigkeitsschätzung mit Deep Learning. In Teil 1 werden wir die physikalischen Grundlagen des Ultraschalls erörtern, um zu verstehen, wie Ultraschallwellen erzeugt werden, sich ausbreiten und in der medizinischen Bildgebung empfangen werden. Wir werden die Grundlagen der Wellengleichung, die Konzepte der Dämpfung, Absorption, Nichtlinearität, Reflexion und Streustatistik behandeln. Anschließend werden wir die Grundprinzipien des Deep Learning und ihre Anwendung in der Ultraschallbildgebung erörtern. In Teil 2 werden wir Methoden zur Parametrisierung und Erzeugung genauer und robuster In-silico-Ultraschallsimulationen diskutieren. Die Fähigkeit, die natürlichen Prozesse der Ultraschallwellenausbreitung und der Streuung im Subwellenlängenbereich in Berechnungsmodellen darzustellen, ist für ein physikalisch informiertes neuronales Netzsystem von entscheidender Bedeutung. Wir zeigen, dass mit Gewebeeigenschaftswerten aus der Literatur ein realistisches In-silico-Phantom als Eingabe für eine numerische Ultraschallsimulation vorbereitet werden kann. Die in dieser Arbeit vorgeschlagene Simulationsmethode ist in der Lage, ein hohes Maß an Realismus zu erreichen und ermöglicht eine Schätzung der Schallgeschwindigkeit. In Teil 3 stellen wir eine neue und neuartige Methode zur Schätzung der Schallgeschwindigkeit in klinischen Brust-Ultraschallbildern vor. Diese Methode baut auf

den physikalischen und Deep-Learning-Grundlagen sowie auf den Simulationstechniken der beiden vorangegangenen Teile auf. Wir zeigen den effektiven Einsatz von tiefen neuronalen Netzen zur Schätzung von Schallgeschwindigkeitskarten sowohl in Phantom- als auch in realen Probandendaten. Diese Arbeit könnte sich erheblich auf die klinischen Ergebnisse auswirken, indem sie die Qualität der nachgelagerten Bilder verbessert und eine quantitative Metrik der Schallgeschwindigkeit hinzufügt, auf der diagnostische Modelle aufgebaut werden können.

# Acknowledgments

This work is the technical summary of five years of personal and professional growth. Along the way, I have had the pleasure to work with many talented researchers from diverse and international backgrounds. I would like to thank my family, especially my parents, Walter and Valerie, for supporting me throughout my studies and my brothers, William and Theodore, for lessening the burden of research with lightheartedness and adventure.

Anyone who knows me will know that during the past five years I have been nearly inseparable from one person, even sharing a desk for large extents. Of course, meeting my wife, Magda Paschali, has been one of the defining events of my time at CAMP and my life. Ever since we met during my first exposure to the CAMP chair at Lake Faak in Austria, she has been a constant source of support and one of the smartest and most hard-working people I have ever met. Her drive and love for science showed me in the darkest times why what we were striving for was important, and she was and still is a constant inspiration.

I would furthermore like to thank Nassir Navab for his scientific and personal guidance and support. This work has been greatly influenced by my collaboration with Jeremy Dahl, who was an excellent scientific advisor for my work on sound speed estimation. Of course, my time in the Interdisciplinary Forschungs Laboratory (IFL), a special place in the Klinikum rechts der Isar in Munich, was defined by the collaborative atmosphere created and curated by Thomas Wendler and Nassir Navab. To all my colleagues in the IFL, thank you for the support, interaction, curiosity, spontaneity, and friendships you have provided and exchanged.

May we all strive to find new and exciting applications of computational methods to diagnose and treat humanity.

# Contents

# Part I

Introduction

# 1 Ultrasound Imaging



## Contents



## 1.1 Introduction

Today, in applications from personal photography, to industrial robotics to perhaps sometime soon widely available autonomous vehicles, we use cameras to capture the world and perceive our surroundings. Ultrasound imaging, which reconstructs an image representing a medium's underlying physical properties, allows one to look within by transmitting and receiving acoustic waves.

While early ultrasound imaging was a tiresome process which at one time required tomographic scanning of patients submerged in a degassed water bath [21, 65], modern ultrasound has come a long way and allows physicians to use small handheld transducers to create in-vivo images.

In contrast to other imaging modalities like CT Scans, ultrasound imaging enables medical diagnosis without ionizing radiation. Furthermore, it is characterized by low cost, ease of use, and wide availability, specifically in comparison to larger and costly MRI and CT scanners.

In the past, ultrasound image quality has been a major drawback of the modality; however, quantitative methods [121] now offer a potential solution by allowing transmitted signals to be adapted based on the tissue being imaged, similar to autofocus in modern digital photography.

To achieve this goal, it is important to tightly integrate previous modeled knowledge of the physics of wave propagation to the design of ultrasound electronics combined with modern machine learning techniques within the medical workflow.

We have recently seen major developments in applications of ultrasound to detect cancers [150], plan and steer needle placement [27, 37, 122], and allow medical robotics to create live adaptive 3D imaging [58, 77]. Computer-aided systems powered by quantitative robotic ultrasound that are able to accurately identify tissue sound speed could significantly reduce the need for biopsies and enable early lesion detection.

This dissertation is built on two main axons. First, we propose a novel pipeline for creating realistic ultrasound simulations of breast tissue to reduce the need for large, expert-annotated datasets. Afterward, we introduce a novel deep learning model capable of accurately predicting tissue sound speed in phantom and in-vivo data after being trained on simulations. Our method is able to bridge the physics fundamentals of ultrasound and the powerful capabilities of deep learning to provide a generalizable and robust sound speed estimation model.

In the following sections, we will discuss some of the underlying physical principles of ultrasound imaging and their interaction with the image reconstruction process. We will cover the basics of acoustic wave physics in Section 1.2, followed in Section 1.3 by a discussion of the specialized hardware that can be used to generate and receive ultrasound waves in human tissue. With this understanding, we will briefly discuss a selection of image reconstruction methods in Section 1.4. Lastly, we will provide an overview of modern clinical applications of ultrasound imaging in Section 1.5.

## 1.2 Physical Principles

By the time an ultrasound image has been formed, the waves that produced that image have undergone various physical transformations on their path through the medium. These include their generation, reflection, refraction, scattering diffraction, and attenuation, as well as advanced non-linear propagation in some cases. Some of these transformations are responsible for clinically viable information in the resulting ultrasound image, while others

can lead to artifacts that deteriorate image quality. In general, every physical transformation a wave undergoes has the potential to be decoded by a receiving transducer to glean information about the tissue through which the wave has passed. To better understand the mechanisms by which such information can be retrieved, we must first review the basic physical principles of the underlying wave physics.

### 1.2.1 Wave Physics

The term wave has many colloquial meanings. We often associate the waves with the back and forth flow of water on a beach or the ripples in an otherwise still pond. Some might think of the transmission of electromagnetic pulses that are responsible for modern telecommunication services and wireless internet in cozy cafés. With a well-rooted knowledge of physics, some might also think of light. Though all of these examples are indeed waves, with similar mathematical descriptions, in the scope of ultrasound imaging, we will focus on the mechanical or acoustic waves.

### 1.2.2 Types of Mechanical Waves

Mechanical waves describe the transmission of energy through a medium via an oscillatory motion of the mediums' underlying particles and the subsequent interaction of the motion of one particle with the motion of the next. Mechanical waves are limited to travel through an elastic solid or fluid by definition and only enable the transmission of energy through a medium and not a net motion of the medium itself. In general mechanical waves can be categorized into two classes: transverse waves and longitudinal waves.

Transverse waves are waves in which the wave motion is perpendicular to the apparent direction of travel of the wave. Our example of a wave on the surface of a pond is a simple example of a transverse wave, since the mechanical offset, or the apparent height of the water relative to the surface, is perpendicular to the direction of travel, outward from the wave source. Transverse waves are generated through the introduction of a shear force to the medium.

The second class of waves is longitudinal waves. Longitudinal waves describe a mechanical motion in the direction of travel of the wave, and particles in the medium oscillate back and forth. Acoustic waves are longitudinal waves through a solid or fluid. In regions where particles have moved towards each other, one speaks of compression or high-pressure, whereas in regions where particles move away from one another, one speaks of rarefaction or low-pressure. Often, the intermolecular oscillation in mechanical waves is explained with a simplified model of masses connected by springs. In a simple mental model, one mass is attached to one spring and in a state of rest connected to a solid and immovable point. In the rest position, the tension in the spring is constant, and the mass is at rest. Should an acceleration be applied to the mass via an oscillatory input, the mass will accelerate. As the mass displaces, the kinetic energy of the mass will be translated into potential energy in the spring. The potential energy of the spring will be translated back into the mass as it is

accelerated back in the direction it came from. Assuming a lossless system, this motion will continue forever.

Our simplified model only consisted of one spring and one mass. To extend this model to a more realistic scenario, we can repeat the system in one direction infinitely many times: an endless row of masses chained together. Given an input force in one direction, the offset in mass will translate through the system, the offset of one mass simulating the offset of the next, and so on. By finally expanding this system in three dimensions, we have a basic mental model for three-dimensional models of wave propagation in three-dimensional space on which we will build upon.

### 1.2.3 Frequency, Sound Speed, Wavelength, Amplitude, Phase

We now have discussed the types of mechanical waves and generated a mental model in which we can observe them. To generalize this model, we must find a way to differentiate different waves. In our mental model of an oscillating system of springs and masses, we can quantify the offset of a given mass from a stationary observer, the time it takes for mass, once set in motion to pass back through its origin. Because this action in our system will repeat indefinitely, we quantify the duration in "times-per-second" [Hz]. This quantity defines the frequency of our wave.

- **Infrasounds**: $f \leq 20Hz$. The human ear cannot perceive acoustic waves in this frequency band. The main application of this frequency band is monitoring for earthquakes.

- **Audible sounds**: $20Hz \leq f \leq 20kHz$. This frequency range describes the hearing range in humans and most animals.

- **Ultrasounds**: $20kHz \leq f \leq 1GHz$. This frequency range contains frequencies higher than the upper audible limit of human perception. Their main application domains are industry and medicine.

Given a single wave propagating through our system, the displacement's speed from one mass to the next defines the speed with which the wave propagates. This speed is a further descriptive factor of our system. In our mental model, with the properties of all springs assumed constant, this speed is also constant. We call this wave propagation speed for mechanical waves, sound speed, or speed of sound. These terms are often used interchangeably. Sound speed is often represented with the variable $c$ in the units [m/s].

To quantify and describe a wave further, we arbitrarily define the points of maximum rarefaction as a trough and the points of maximum compression as peaks. This naming convention may seem strange for longitudinal waves but is, in fact, borrowed from transverse waves, where peaks and troughs are the highest and lowest physical points on a wave, respectively. In longitudinal waves, these extrema represent the maximum and minimum *pressure* within the wave. Suppose we track the pressure at a given point in the medium. In that case, we will

get an oscillating function that can be represented with a generalized combination of sin and cosine functions. The value of the function with respect to the mean is called the amplitude.

As the waves we have been examining pass through the medium, they transport energy measured in Joules [J]. The transportation speed of energy through an area defines the power of the wave in Watts [W].

$$\text{Wave Power } P = \frac{Ap^2}{\rho c} \text{ for perpendicular wave propagation direction}$$

Here $\rho$ is the medium density, $p$ is the sound pressure, $A$ is the area of integration, and $c$ is the sound speed of the medium. When this power is integrated over an area in our three-dimensional model, we call that property the intensity of the wave [W / $m^2$].

$$\text{Wave Intensity } I = \frac{P}{A}$$

Given a wave frequency $f$ and a sound speed of medium $c$, one can empirically derive that a wave propagating through the medium has a distance between one peak and the next.

$$\text{Wavelength } \lambda = \frac{c}{f}$$

The sound speed of a wave is determined by the material through which the wave travels, and the frequency of the sound wave is dictated by the source which produced the wave. Therefore wavelength is source- and medium-dependent.

As an observer, we can watch waves pass by, traveling at a sound speed $c$ and with frequency $f$, but how can we quantify which "part" or phase of the wave is at our position at any given point in time? The tracking of cyclical wave progression over time is strongly related to circular motion and is therefore often described by an angle in degrees or radians. As a matter of definition, the peak of a wave defines $0°$, and a static observer would notice the wave pass through $0°$ once every $\frac{1}{f}$ seconds. All intermediate positions of the wave in position and relative velocity are defined by the phase of the wave between $0°$ and $360°$. A representative diagram can be seen in Figure 1.1.

Often in ultrasound imaging, longitudinal waves are transmitted through biological tissue. Until now, we have only spoken about how waves propagate through a medium and their relative mechanical offset to their transmission direction. In order to be able to make an image, we have to understand how the waves we send out into the tissue can return to the point of transmission, be registered, and be attributed to a point in space.

### 1.2.4 Reflection

Until now, we have discussed the propagation of waves in a simplified homogeneous model. When an acoustic wave encounters a boundary between two media, the behavior of the

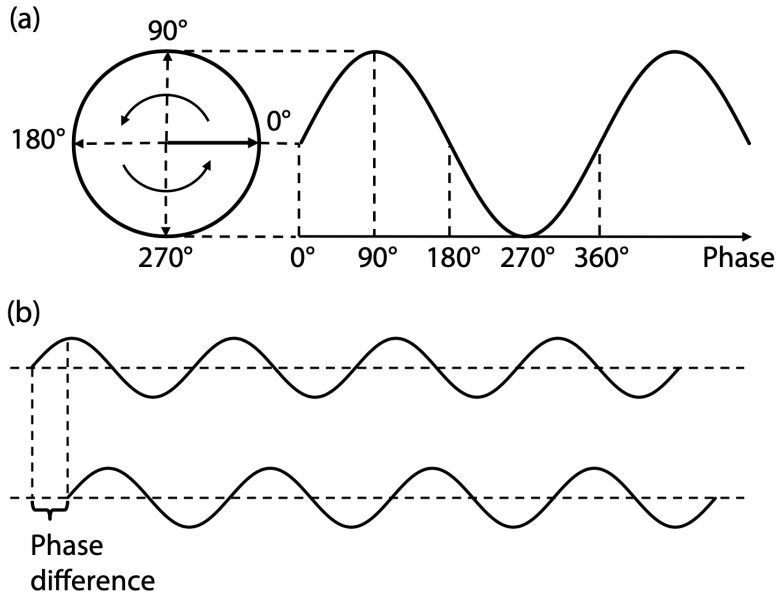



**Fig. 1.1.** The phase of a wave is defined by the cyclical relationship of an oscillating wave, which can be described by a rotating phasor, or angle and magnitude, on a circle. Subplot (a) depicts this relationship. A phase offset is defined by a shift in the relative phase of two waves, commonly referred to as phase difference. An example is displayed in subplot (b). [64]

wave at the boundary is dependent on the physical properties of the two media, respectively. Depending on the physical properties of the media, some energy can be reflected back, and the remaining energy continues through the new media. The action of reflection is a key property in pulse-echo ultrasound, and we will briefly discuss the physical properties that lead to reflection.

The acoustic impedance of a medium is defined as:

$$z = \frac{p}{v}, \qquad \text{(Specific Acoustic Impedance)}$$

where $p$ is the local pressure in the medium and $v$ is the local particle velocity. This is analogous to Ohm's Law which describes electrical impedance as the ratio of the electrical voltage to electrical current. A secondary formulation of acoustic impedance can also be derived [86], which defines acoustic impedance in terms of medium sound speed $c$ and density $\rho$:

$$z = \rho c. \qquad \text{(Characteristic Acoustic Impedance)}$$

Since this formulation of acoustic impedance is based on macroscopic properties, it is referred to as the characteristic acoustic impedance.

With the concept of acoustic impedance defined, we are ready to explore the concept of reflection. Given two media, medium one and medium two, each with a separate acoustic impedance, we can evaluate the wave propagation over the boundary between medium one and medium two. At the boundary, there are two possible outcomes for a propagating

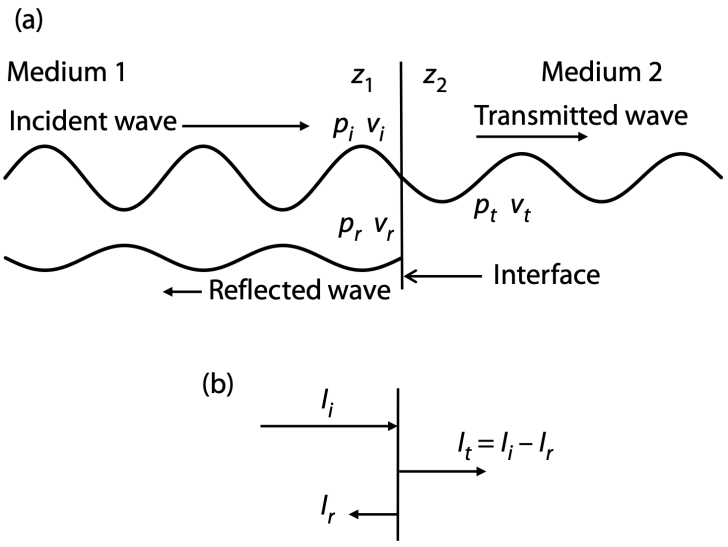



**Fig. 1.2.** The diagram above shows how wave reflection occurs. The total particle velocity and pressure at every point must be contiguous, even where there is a change in acoustic impedance. This results in the reflection of a portion of the wavefront back in the direction of the sender, as can be seen in (a). Since the total intensity must remain constant, the intensity of the impinging wave can be written as the sum of the reflected and propagated waves. This property is depicted in (b) [64].

wavefront. The first is a transmission of the wave from medium one into medium two. The second is a reflection of the wave at the boundary back through medium one in the opposite direction. In reality, both outcomes occur in proportion to the difference in acoustic impedance between mediums one and two. When the change in acoustic impedance is large, much of the energy is transmitted back through medium one in the opposite direction, and very little is transmitted through medium two. Conversely, when the change in acoustic impedance is low, much of the energy continues on through medium two, and little is reflected back through medium one. This phenomenon occurs in order to keep the total particle pressure and velocity at a micro-level constant. Given a change in tissue properties, it must hold that:

$$p_t = p_i + p_r \tag{1.1}$$

$$v_t = v_i + v_r \tag{1.2}$$

This means that the total particle pressure and velocity, i.e., the sum of the incidental wave-particle pressure and velocity ($p_i$, $v_i$) and reflected wave-particle pressure and velocity ($p_r$, $v_r$) in a given continuum must equal that of the transmitted particle pressure and velocity. With this formulation and the definition of Specific Acoustic Impedance, we can derive the fact that:

$$z_1 = \frac{p_i}{v_i} = \frac{p_r}{v_r},$$

and

$$z_2 = \frac{p_t}{v_t}.$$

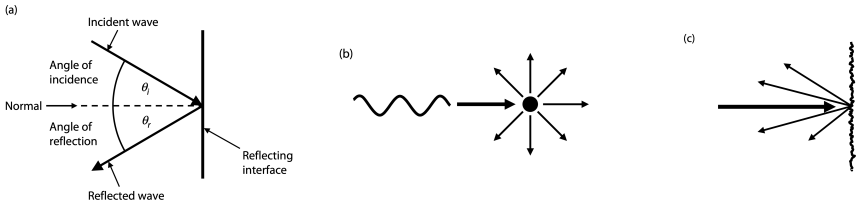



**Fig. 1.3.** The behavior of an ultrasound wave with an object depends on the relative size and shape of the object to the wavelength of the wave in question. Given a relatively large and flat surface, one can expect reflection of the wave in proportion to the angle of incident of the wave $\theta$ as can be seen in (a). Given a circular, sub-wavelength object, often referred to as a scatterer, the wavefront is reflected (scattered) in all directions after interaction, as can be seen in (b). Similarly, a rough surface, where the geometry of the surface is smaller than the wavelength of the propagating wave, reflects the wave in multiple directions back towards the sender, similarly to a point scatterer (c.f. Section 1.2.7) [64].

From these two equations, we can formulate the reflection ratio $R_A$ of reflected and transmitted pressures at a perfect interface of mediums one and two:

$$R_A = \frac{p_r}{p_i} = \frac{z_2 - z_1}{z_2 + z_1}$$

This ratio of the reflected pressures defines the amplitude of the reflected wave and, as we will see later, the intensity of the interface on the reconstructed ultrasound image. We can further describe this formulation as the intensity ratio $R_I$, since we know that the intensity is proportional to pressure squared, i.e.:

$$\frac{I_r}{I_i} = R_I = R_A^2 = \left(\frac{z_2 - z_1}{z_2 + z_1}\right)^2.$$

As we previously discussed, the intensity of a wave is a measure of the rate of energy flow (power) through an area. At any given interface, this power must be conserved as the wave "splits" into the transmitted and reflected waves. This means that

$$I_t = I_i + I_r,$$

and therefore the transmission coefficient $T_i$ can be defined as:

$$T_i = \frac{I_t}{I_i}.$$

Up until this point, we have examined a perfectly orthogonal transmission path through a hypothetical interface. Of course, wave transmission is not always so simple. Assuming perfect reflection, let us examine the wave behavior in two dimensions by adding an angle of incidence $\theta_i$, i.e., a non-orthogonal wave impingement case.

For a flat and smooth surface, the fully reflected wave will reflect at an angle of reflection $\theta_r$ of the same magnitude as $\theta_i$ but will be reflected across the surface normal. A simple diagram of this property can be seen in Figure 1.3 a). Moving forward, we will examine some of the nuances and complexity that can occur in wave propagation, how they can be modeled, and their consequences.

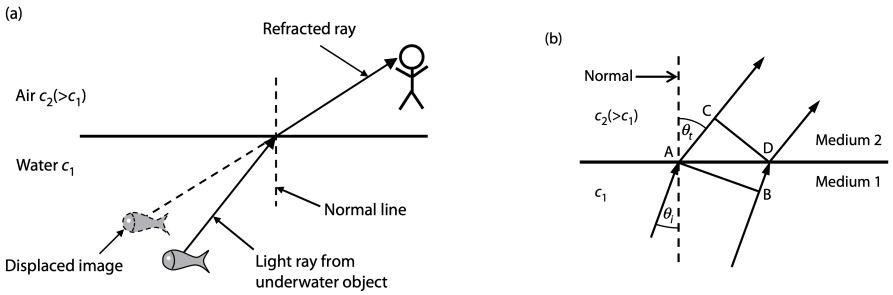



**Fig. 1.4.** Refraction describes the bending of a wave as it travels from material one with one set of physical properties to another material (material two). The above example in subplot (a) shows a refracting ray of light as it passes from air into water. The same phenomenon is apparent in acoustic waves and can be responsible for the shifting of objects in ultrasound images. Subplot (b) shows the mechanics of the phenomenon on a smaller scale by imagining the ray of light has a non-neglectable width. With the constraint that both rays must remain parallel, ray A passes into the new medium first and begins to travel faster than ray B. This leads to a rotation in the overall propagation direction of the wavefront due to the aforementioned parallelism constraint. Once both waves are in the new medium, the overall travel direction of the wavefront has changed [64].

### 1.2.5 Refraction

Until now, we have discussed orthogonal reflection with a proportional transmission term $R_T$ which is dependent on the acoustic impedance of media one and two. We have also briefly discussed the full reflection case for an incoming wavefront at a non-zero angle of incidence $\theta_i$ and the subsequent angled reflection $\theta_r$. In this section, we will further discuss the case of angled incidence but add the level of complexity of varying sound speeds and therefore varying acoustic impedance in media one and two (c.f. Equation Characteristic Acoustic Impedance). In the case of varying sound speed, the physical phenomenon of *refraction* comes into play.

Refraction is the phenomenon of the change of direction in wave propagation as a wave passes between two media with a change in sound speed magnitude. Refraction can be commonly seen in lightwave propagation when viewing an object underwater. Due to the varying material properties of light and water, the object being viewed appears to be in a different location than it is in reality due to refraction. The concept of refraction is often modeled with one-dimensional ray diagrams as seen in Figure 1.4 a). These diagrams are useful in modeling the concept, but a deeper level of understanding of the underlying mechanics can be gleaned by a two-dimensional wave propagation example as can be seen in Figure 1.4 b).

Here, the two-dimensional wave is represented with two one-dimensional rays traveling parallel to one another. Both before and after the interface, the wave rays must be traveling parallel to one another. As the wavefront approaches the interface at the angle of incidence $\theta_i$, ray A intercepts the medium boundary first. In the new medium, ray A travels more quickly relative to the propagation speed of ray B. Since both rays must travel parallel to one another while in the same medium, the change in medium sound speed leads to the relative rotation of the wave propagation angle. By the time ray B has entered medium two, the parallel rays

are now traveling at a new angle $\theta_t$. The relationship between sound speeds $c_1$ and $c_2$ and the angles of incidence and transmission $\theta_i$ and $\theta_t$ can be described by Snell's Law:

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{c_1}{c_2}. \qquad \text{(Snell's Law)}$$

The proportionality described by Snell's law dictates that the transmission direction remains the same for two media with constant sound speed $c$. When the sound speed of $c_2$ increases, the angle of transmission $\theta_t$ increases proportionally. Should the sound speed $c_2$ decrease relative to $c_1$ the angle $\theta_t$ decreases. We will see later that refraction is one of the major factors leading to imaging errors in medical ultrasound images with assumed constant sound speed.

### 1.2.6 Non-linearity

Until now, we have been working under the assumption that materials have intrinsic properties, e.g., sound speed, that affect the propagation of a wave in a given material. Furthermore, we have discussed the linear relationship between the amplitude waves at their source and elsewhere along their propagation path. This model is called linear wave propagation and works well to describe wave propagation for relatively low amplitudes. The concept of non-linearity describes the breakdown of these linear relationships in wave propagation's for large waveforms with high-pressure amplitudes (e.g.,>1 MPa) [36, 67]. With the large amplitude discrepancies, local medium properties begin to deviate from mean tissue properties. Higher pressures (compression) regions of the wave peaks lead to higher local sound speeds in the peaks. Low-pressure regions (rarefaction) lead to lower local sound speeds in the troughs. These local sound speed discrepancies lead to increased and decreased local phase speed of the wave. This, in turn, leads to the peaks of the wave "over taking" the troughs of the wave as the wave progresses through a given medium. This process can lead to a sinusoidal wave transforming into a sawtooth wave as it progresses. This non-linearity has effects on the pulse spectrum tissue response and resulting images.

### 1.2.7 Scattering

Until now, the scale of the interfaces we have examined, though not explicitly stated, has been much larger than the wavelength of the propagating wave. When a wave interacts with an interface that is much smaller than the wavelength, the physics of the resulting interactions are no longer covered by the simplified models previously presented. The resulting phenomenon is called scattering, and we will briefly discuss the underlying physics of the phenomenon and how it can be modeled.

The power of a scattered signal is proportional to the size $d$ of the scattering target and the wavelength $\lambda$. For targets that are orders of magnitude smaller than the wavelength ($d \gg \lambda$, the power of the scattered signal $W_s$ is proportional to the sixth power of the scatter size $d$

and inversely proportional to the fourth power of the wavelength $\lambda$. This is referred to as Rayleigh scattering.

$$W_s \propto \frac{d^6}{\lambda^4} \propto d^6 f^4 \tag{1.3}$$

Scattering is very relevant in medical ultrasound images due to the relatively large wavelength of the sound waves propagating through the tissue and the multiple scales of media interfaces in biological tissue. The primary scale of the scatterers in medical ultrasound imaging is still not well understood [152]. It has been hypothesized that scattering is a result of cell walls, cell nuclei, and protein structures, but until now, the author is unaware of a general consensus.

Regardless of the causal origins of scatters in medical ultrasound, the existence of scattering in medical ultrasound is clear. The result of scattering in B-mode image reconstructions is speckle or the statistical constructive and destructive interference (c.f. Section 1.2.11) of signals caused by sub-wavelength scattering. Speckle appears as a texture in a B-mode image. Speckle can vary widely based on the organs being imaged in medical ultrasound and can be an indicating factor in many diagnoses. For example, in the diagnosis of fatty liver disease, speckle appearance is often a contributing factor to the final diagnosis [79, 140].

Speckle is commonly modeled as a random walk [128], which models the random sub-wavelength energy reflection between multiple points like points or scatterer in the domain. These sub-wavelength scatters referred to as **diffuse scatterers** due to their unknown reflection coefficient and position [33]. The duration of the random walk is defined by the sampling frequency of the transducer. The number of steps of the random walk is derived by the number of random scatterers the wave encounters during this duration. The volume covered by the wave during this period is called the **isochronus volume** [33, 105] (c.f. Section 1.4). The vector sum of a random walk within the isochronous volume can be defined as:

$$ae^{j\psi} = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} a_k e_k^{j\psi_k}$$

In the radio frequency (RF) domain or the domain of a complex waveform consisting of the base frequency and layered reflection signals, the first-order statistics of the speckle is zero-mean with a Gaussian distribution [33]. It can also be shown that the amplitude of speckle is Rayleigh distributed in the envelope detected domain (c.f. Section 1.4 for a given envelope detected signal, while the phase has a uniform distribution between $-\pi$ and $\pi$ [33]. This leads to the fact that we can define the signal-to-noise ratio (SNR) of speckle.[1] A Rayleigh distribution is defined by:

$$\mu_V = \sqrt{\frac{\pi}{2}\sigma_A} \tag{1.4}$$

$$\sigma_V^2 = \left(2 - \frac{\pi}{2}\right)\sigma_A^2 \tag{1.5}$$

[1] A formal definition of SNR will be presented in Section 1.4.4

$$R_A(x,y) = \underbrace{\mu_V^2}_{\text{medium reflectively scaling factor}} \underbrace{g(-x,z) \otimes g^*(x,z)}_{\text{point spread function}} \tag{1.6}$$

The SNR of an arbitrary patch of speckle can therefore be written as:

$$\text{SNR}_0 = \frac{\mu_V}{\sigma_V} = \sqrt{\frac{\frac{\pi}{2}\sigma_A^2}{\left(2-\frac{\pi}{2}\right)\sigma_A^2}} = \sqrt{\frac{1}{\frac{4}{\pi}-1}} = 1.91.$$

From this, we can see that speckle is multiplicative and standard deviation scales with the amplitude of the underlying mean signal. Furthermore, the SNR of speckle is *always* 1.91 regardless of focus, frequency, aperture etc [182].

This model of speckle statistics is valid for a large number of diffuse scatterers in a volume, given that a random walk tends towards infinity. It is said, though, that at minimum, ten scatterers per resolution cell are needed to achieve Rayleigh statistics [33, 182].

For second-order statistics of speckle, i.e., temporal and spatial coherence, it can be said that speckle has a high, near-constant temporal coherence and a more complicated spatial coherence. For a given transmission pattern and position, the speckle pattern generated in the resulting B-mode image does not change [33, 50]. This points us in the direction of saying that for a given transmission, the wave propagation path is deterministic. Mathematically, the temporal coherence can be written as $R(\tau) = 1, \forall\, \tau$. In reality, keeping all variables constant between transmissions is near impossible in a clinical setting, but nevertheless, one can say that the temporal coherence of speckle is very large.

The size of the speckle is directly related to the size for full-width half max (FWHM) 1.14 (c.f. Section 1.4.4) of the auto-correlation function and is in an indicator of the resolution of a given ultrasound machine [50, 92, 106].

The study of speckle statistics is a large field of research, the scope of which is far beyond the minor introduction in this work. The basic overview presented here should allow the reader to see how the topic of speckle applies to the topics of the following sections, but there are a complete resource on statistical optics for those readers who are interested [50].

### 1.2.8 Attenuation

A further physical property of acoustic wave propagation that is very important in medical ultrasound imaging is attenuation. Attenuation represents the reduction of energy in the wave as the wave propagates through a medium. This reduction is often characterized by an exponential decay in wave intensity over the distance traveled by the wave.

Because energy can neither be created nor destroyed, there are many mechanisms by which the wave loses energy as it travels through the medium. One such mechanism, which only contributes a small amount to the attenuation of ultrasound waves through biological tissue, is

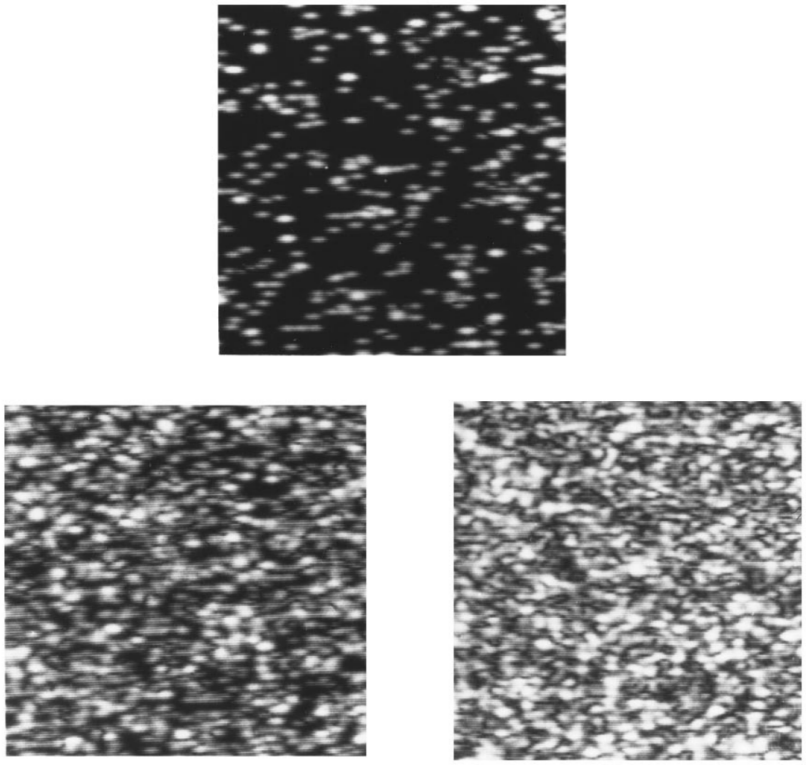


Fig. 1.5. Scatters, the underlying source of speckle noise, can be simulated. Scatter density has an influence on the statistics of the returning signal. A fully fledged speckle is said to be achieved when a scatterer density of 10 scatters per wavelength cell has been achieved. Above are three examples of 1 2, and 6 scatters per wavelength [185]. Reprinted with permission from Keith A. Wear, Robert F. Wagner, David G. Brown, Statistical properties of estimates of signal-to-noise ratio and the number of scatterers per resolution cell . ©1997, Acoustic Society of America.

scattering, as we will discuss in Section 1.2.7. This makes sense since the energy that has been "scattered" is no longer considered to belong to the intensity of the wave that is transmitted on through the tissue.

The largest contributing factor of attenuation in medical ultrasound imaging is absorption.

### 1.2.9 Absorption

Absorption describes the translation of kinetic energy in the wave to thermal energy in the medium by the physical mechanism of friction. As a wave passes through a medium, the mechanical deformation of the medium is not completely lossless. Every deformation results in mechanical motion along molecular or cellular boundaries. This motion results in the translation of mechanical energy into thermal energy through friction. This thermal energy remains in the medium and reduces the wave intensity during its further propagation.

Since the rate of absorption is dependent on the number of mechanical movements undergone by the medium frequency of the wave, it follows that absorption is a frequency-dependent

property. As the frequency of a wave passing through a medium is increased, the rate of attenuation of the signal by absorption also increases.

### 1.2.10 Ultrasound Beams

Until now, we have predominantly focused on a ray approximation of wave propagation, with a minor two-dimensional expansion of that model with a planar wave for our discussion of refraction in Section 1.2.5. In reality, waves are three-dimensional. We know from our understanding of the expanding wavefront on a pond's surface after a pebble has been cast into the pond that waves also experience concentric expansion given a point-like impulse in space.

In this section, we will discuss how we can control and manipulate the propagation direction of waves by controlling the layout and timing of the generating source. A special wavefront that has been created in such a way as to propagate mainly in one direction along a narrow corridor is called a beam. Subsequently, the art of generating such special waves, i.e., the art of generating beams, is referred to as beamforming and is the subject of this section.

In order to understand how beamforming can take place, we must first understand the interactions that can occur between waves when waves from separate sources cross each other spatially along their propagation path in a medium.

### 1.2.11 Interference

Until now, we have only observed individual waves. The concept of beamforming is fundamentally based on the complex interactions of multiple waves in a medium. In order to simplify these complex interactions, we will reduce our model to the simple interactions of two waves overlapping spatially in the one-dimensional case.

Put simply, when two waves overlap spatially, as is pictured in Figure 1.6, their amplitudes are added. When the two positive amplitudes overlap, one speaks of *constructive interference*, and the resulting amplitude equals the sum of the two overlapping waves. On the other hand, when a positive and a negative amplitude overlap spatially, one speaks of **destructive interference**, and the resulting absolute amplitude decreases. Again here, the resulting amplitude equals the sum of the two overlapping wave amplitudes.

### 1.2.12 Diffraction

We have been bouncing back and forth in our two-dimensional wave models between a wavefront that progresses in one direction with a lateral width, often referred to as a "plane-wave" due to the planar appearance, and our model of a concentric wave, propagating out from a point source like a stone on the surface of a pond. But what dictates the form and propagation properties of any given wave? In reality, the properties of wave propagation are dependent on the relationship between the size and geometry of the source generating the wave, often called the *aperture* and the wavelength of the wave. If the aperture is smaller than

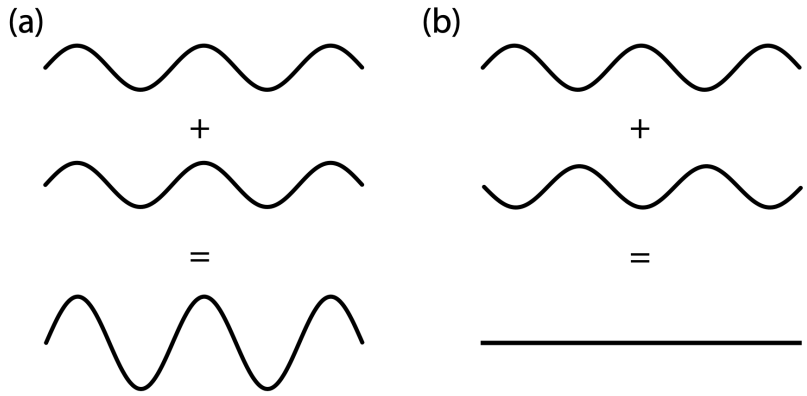



**Fig. 1.6.** Above simple is an example of wave interference given one-dimensional waves. Waves that have the same phase interfere constructively, and the resulting amplitude is double the input amplitudes. Wave with opposite phase will result and destructive interference, and the amplitudes subtract from one another. In the above case, the amplitudes have equal magnitudes and are therefore canceled out completely [64].

the wavelength, the wave spreads from the source and diverges, like in the case of a small pebble being cast into a pond. This effect is called **diffraction**. Conversely, a large and flat paddle, like those often found in wave pools, is proportionally closer in relationship to the wavelength of the waves it creates, and therefore the waves that propagate perpendicular from the source travel as planar fronts.

We can define a relationship between the large and small source cases by creating a system of small, point-like sources in a line, as can be seen in Figure 1.7. The interference of the diverging waves of the same frequency from the point sources leads to the approximation of a plane wave as the waves propagate. The wave sections propagating in the same direction interfere constructively, while those sections not propagating in the same direction often interfere destructively. By adjusting the spatial distribution of smaller apertures, larger apertures can be approximated. This spatial layout of an aperture can also be electronically simulated by manipulating the timing of the transmission of the waves from each point; a method often referred to as active scanning or steering.

By leveraging the properties of interference and diffraction, we will see in the subsequent sections how waves can be generated in order to converge to a point, diverge forever, or propagate as a plane, all depending on the spatial layout and temporal arrangement of the wave sources. The application of beamforming has a multitude of applications in and beyond medical imaging, for example, in satellite and radio communications, seismology, and radio astronomy.

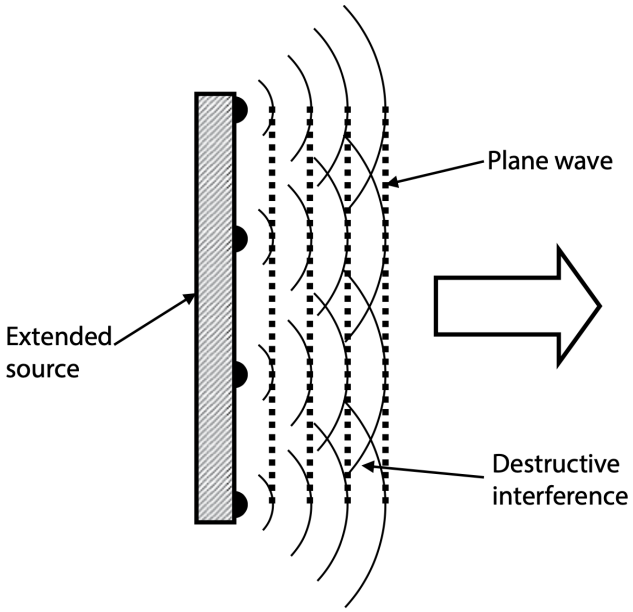



**Fig. 1.7.** The individual waves from an array of point sources can interact constructively and destructively. Over the propagation path, the individual wavefronts create a larger coherent wavefront. The shape of the wavefront depends on the geometry of the array. The above example shows a linear array creating a plane wave [64].

## 1.3 Ultrasound Hardware

We have discussed the physics of wave propagation once a wave is in a medium. We have briefly discussed the relationship an aperture's shape has with the propagation pattern of a wave and how waves can be arbitrarily laid out and steered based on a controlled generation from smaller point-like sources. In this section, we will get more practical and discuss the devices that can both transmit and receive ultrasound wave called medical transducers and their application-dependent form.

An ultrasound transducer is the practical embodiment of a device that is able to interact with all the physical phenomena we have discussed up until now. Medical ultrasound transducers specifically commonly consist of an array of discrete piezo-electric elements which can generate and sense acoustic waves in a medium with which they are in physical contact. Piezo-electric elements are a special combination of piezoelectric materials, such as quartz which react to electrical voltage with physical expansion or contraction and conversely react to physical expansion and contraction by generating an electrical potential. This useful property lends itself to construction in ultrasound transducers.

The arrays of piezoelectric material are commonly mounted between a backing plate and a lens, which transfers the physical energy from the hard elements into soft biological tissue. The arrays are linked to individual electrical leads that oftentimes run through a handle and a cable to a separate piece of hardware for signal generation and processing. An overview of a typical transducer layout can be seen in Figure 1.9.

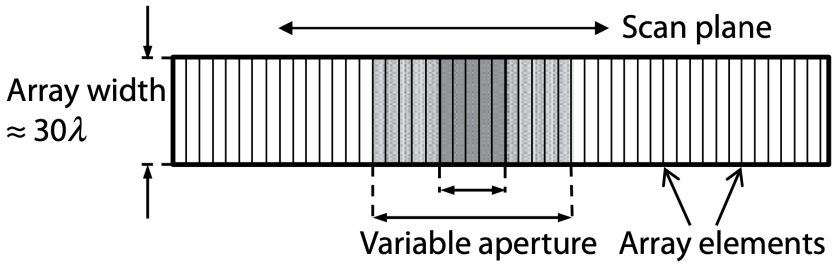



**Fig. 1.8.** An example of a linear array can be seen above, with rectangular elements. The elevation of elements is often around $30\lambda$. The total width of the transducer is referred to as the aperture. A subset of elements can be activated depending on the desired transmission region. This subsection is called the sub-aperture [64].

Medical ultrasound transducers can have diverse element layouts based on the medical application for which they are designed. For example, for many superficial organs of interest, such as mammaries, thyroid, and vasculature, a linear array transducer is often used. Linear array transducers generate a rectangular field of view of the tissue.

Curvilinear arrays place the elements of the transducer on the arc. The waveforms generated from a curvilinear transducer, therefore, are transmitted with a wider field of view which becomes wider with depth. For this reason, curvilinear transducers are often used for abdominal applications, where a wide field of view allows one to visualize larger abdominal organs. Phased-array transducers are similar to linear arrays but have their elements much closer together. They transmit a trapezoidal window which, like curvilinear transducer arrays, gets larger with depth. Due to the small aperture and proportionally large imaging window, phased array transducers are often used in cardiac imaging, where the transducer must be placed between ribs for proper imaging.

Endocavitary transducers consist of small arrays mounted to an elongated probe. These transducers allow imaging from within human cavities such as the rectum or the vagina and are used in both the fields of obstetrics and urology for imaging and diagnostics. Intravenous transducers are small transducers placed on the end of a catheter and can image the interior of vasculature and even the human heart.

Lastly, two-dimensional matrix array probes are relatively new developments in ultrasound imaging and allow piezoelectric arrays to be created as a two-dimensional matrix. By transmitting from this two-dimensional matrix, three-dimensional images can be created. These three-dimensional ultrasound images can currently be used for obstetric prenatal imaging and trans-cranial imaging of the brain for navigation during surgery.

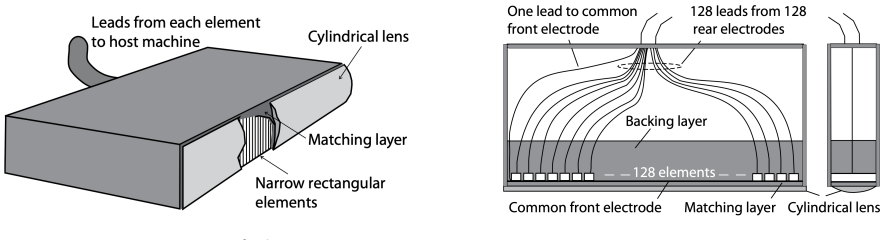



**Fig. 1.9.** On the left, we show a cut-away view of a linear transducer with a cylindrical lens, a matching layer, and the linear array elements. The matching layer helps alleviate the significant material differences between the hard piezoelectric elements and the properties of human tissue, while the lens focuses on wavefront in the elevational plane. On the right, we show a typical cross-sectional layout of a transducer showing wires, also called channels, leading to the piezoelectric elements, which are mounted on an acoustic backing material and covered by a perfect matching layer and lens. The acoustic backing material ensures that the power from the piezoelectric elements moves forward out of the transducer for imaging purposes and not backward into the transducer [64].

## 1.4 Image Generation

Ultrasound imaging is based upon the pulse-echo principle of acoustic waves. We have extensively discussed the physics of wave propagation in Section 1.2. Now we will visit the more practical implications of these physical principles and how ultrasound images are generated through the interaction of a medical ultrasound transducer, wave propagating medium of interest.

This section will cover the concept of a pulse in ultrasound imaging. Then we will explore the point spread function (PSF), a concept that describes a medical ultrasound device's axial and lateral resolution. Lastly, we will briefly discuss the idea of fractional bandwidth of a transducer and standard ultrasound imaging frequencies.

### 1.4.1 Transmission Methods

Though theoretically, the transmitted signal for ultrasound imaging is arbitrary, practically, often sinusoidal pulses are used as pictured in 1.10. A sinusoidal pulse is an amplitude-modulated Gaussian envelope over a carrier signal of frequency $f$. Common transmission frequencies in medical ultrasound imaging range between 2 and 18 MHz [24]. The axial resolution of a transmitted pulse is half the pulse duration [33]. For this reason, to increase imaging resolution, a short pulse is desirable. For higher frequency carrier frequencies, shorter pulse durations are possible. This comes with the trade-off of shallower imaging depth due to frequency-dependent attenuation as discussed in Section 1.2.8.

The PSF is defined as the response of an ultrasound transducer when an individual sub-wavelength scatterer is imaged Figure [29]. The size and shape of a PSF encode all information about the ultrasound imaging system and are comparable to the impulse response of a 1D time-domain system [33]. The pulse form is embedded on the axial cross-section of the PSF. The resolution of an imaging device is half the pulse width. Laterally, the cross-section PSF

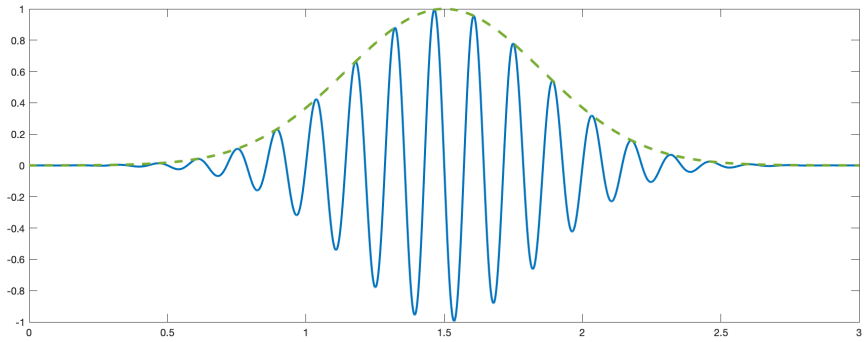



**Fig. 1.10.** Here, we show an example of a sinusoidal pulsed wave with a Gaussian envelope. This waveform is typical in ultrasound imaging. In general, the shorter the pulse, the higher the resulting image's axial resolution. Wave duration is measured in cycles, i.e., how many full waveforms can fit in the pulse duration. A common measure of duration is 1-3 cycles depending on the application.

encodes the lateral resolution of the imaging device. More information on the relationship between aperture, beam shape, and lateral resolution can be collected in [29, 33].

A further way of describing an ultrasound transducer is via its fractional bandwidth, i.e., the spectrum of frequencies sensed by a transducer. A wider bandwidth enables the use of shorter ultrasound pulses and therefore improved axial resolution [64]. The fractional bandwidth (FBW) of a transducer is defined as the width of the full-width half-max (FWHM) measurement relative to the transmit frequency. In medical ultrasound imaging, modern transducers often achieve FBWs of over 70% [109].

The **isochronous volume** refers to the volume of space that the wave occupies at a given point in time [44]. This volume $v(t, m)$ is dependent on the propagation time and the sound speed and waveform transformation the wave experiences as it propagates through the medium $m$. The larger the isochronous volume, the larger the spatial distribution scatterers that influence the resulting image contrast at a given point in time $t$. Therefore, the size and layout of the isochronous volume is a further factor in the understanding of image quality in medical ultrasound imaging.

### 1.4.2 Reconstruction and Beamforming Techniques

When a medical ultrasound receives pulse-echo tissue signals, they are not yet in an interpretable representation but rather a collection of amplitude and frequency modulated signals upon a carrier signal. There is a multitude of steps that these signals must undergo both before transmission and after reception to generate a medical relevant and interpretable image.

The complete imaging pipeline can be seen in Figure 1.11. A representative ultrasound system consists of a discrete set of components, including a transmitter, a transmit-receive switch, a time gain compensation (TGC) module, a beamformer, a signal processing module, a scan-conversion module, and a post-processing module. In the following passages, we will focus mainly on the beamforming component in the context of the complete hardware set.

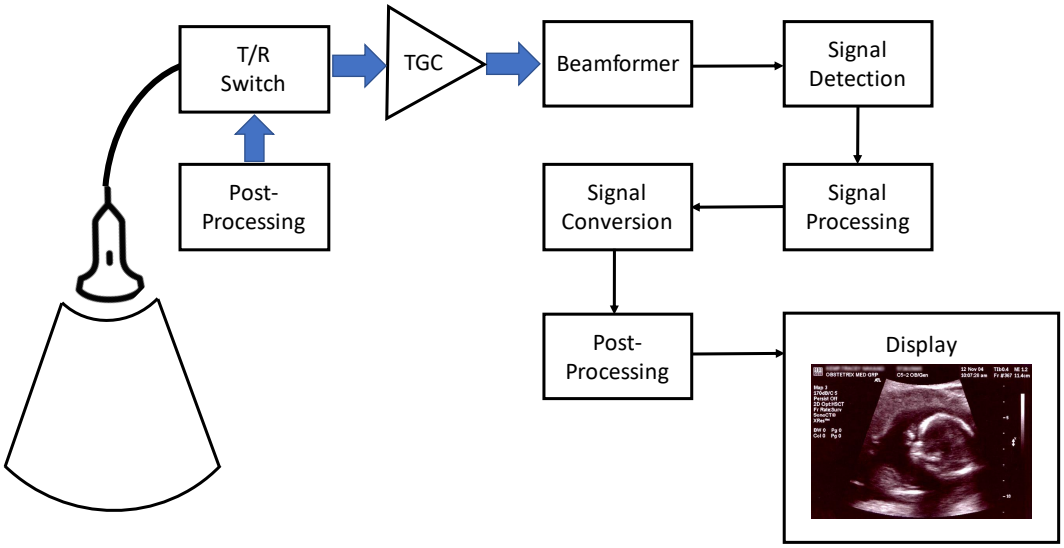



**Fig. 1.11.** Overview of the primary imaging steps of an ultrasound image. A switch allows for both transmit and receive settings. Received signals are processed via TGC to compensate for attenuation in the medium, be amplifying signals progressively over their depth of origin. The beamformer step delays and combines received signals by allocating a signal received in time to a location in a two-dimensional space. Further processing and filtering are then performed before the spatial signals are scaled to pixels on the device's screen in the scan conversion step. Once an image is created, post-processing in the image domain can be performed, and the image can be displayed.

Briefly, the transmitter, often made of a piezoelectric array, is connected to the beamformer module via a transmit and receive switch. The beamformer delays transmitted and received signals for both transmit and receive beamforming. The mode of the transmitter switches from transmit to receive and vice versa with the T/R switch. When the receive signals have been appropriately delayed by the beamforming module, the "signal detection" module performs envelope detection to separate the "tissue signal" from the carrier signal, often via the Hilbert transform. Lastly, scan conversion is performed, by which temporal signals are assigned a spatial location the 2D grid of the B-mode image, and post-processing is performed to filter artifacts or apply clinically relevant filters to the final image [33, 64].

### 1.4.3 Transmission Techniques

The process of *steering* and *focusing* acoustic wave-fronts both before transmission and after reception is called beamforming, getting its name from the radar field [33, 64]. Beamforming improves the directionality, sensitivity, resolution, and SNR of an ultrasound system.

When developing the aperture of an ultrasound transducer, an aperture that creates a narrow beamwidth is desirable for higher resolution, and thereby the ability to distinguish points that are close together (c.f. Section 1.2.12). The ability to focus a wavefront reduces the beamwidth at focal depths and allows for a tighter beam pattern than possible with only the transmit aperture. For steering and focusing of an array element, each element is treated as a point source in accordance with Huygen's Principle [68]. The resulting field of constructive and destructive interference results in a global wavefront built upon the signals of the point sources (see Figure 1.12. Elements of the transmitting array are activated or fired in a

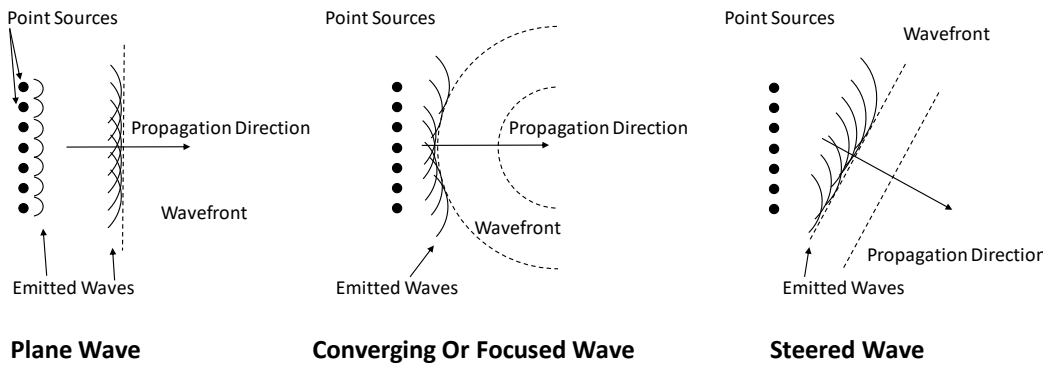



**Fig. 1.12.** Overview of an exemplary diagram of focus and steering mechanisms. On the left, a plane wave is generated via the transmission of multiple point sources simultaneously. In the middle, transmit focusing is applied, which transmits the outer elements first and the inner elements successively later to generate a focal region of constructive interference. On the right, a plane wave is shown to be steered at a constant transmission angle via transmit delays of individual elements. All of these mechanisms can be parameterized and combined in modern ultrasound imaging.

choreographed fashion such that the total propagation is focused on a focal point in the medium. The timing of the element firing is dependent on a set of firing delays $\{\tau_0 ... \tau_n\}$ for an array of $n$ elements.

Based on this timing, a focal point in the medium can be defined at which the maximum constructive interference is achieved, leading to a strong reflection of the material at that point in space. This carries the advantage that peripheral reflections are proportionally less intense, thereby increasing the signal quality of the returning wave. Similarly, linear delays across the aperture result in planar waves with a constant steering angle $\theta$ proportional to the delay between neighboring elements. This method of transmission is equivalent to setting the focal point to be infinitely deep in the medium. This method of transmission was first proposed for imaging by [112], who showed that coherent compounding of transmitted plane wave transmissions and varying angles can lead to faster image acquisitions than with focused waves, the gold standard up until then. Lastly, setting the focal point behind the transmitting array can create diverging waves. Diverging waves have been recently shown to be beneficial for cardiovascular imaging due to their higher temporal resolution [168].

Despite modern imaging technologies, ultrasound images still suffer from imaging artifacts due to medium property inhomogeneities and the current lack of methods for their compensation. Sound speed fluctuations in the medium are responsible for reverberation, which reduces the focus of the resulting ultrasound image. Reverb can present itself in two forms. Gross reverb occurs when an ultrasound beam is reflected multiple times between two often parallel interfaces before returning to the transducer, leading to the characteristic “echo” beneath the interface in the image. Local reverb is created by the same method but on a smaller scale

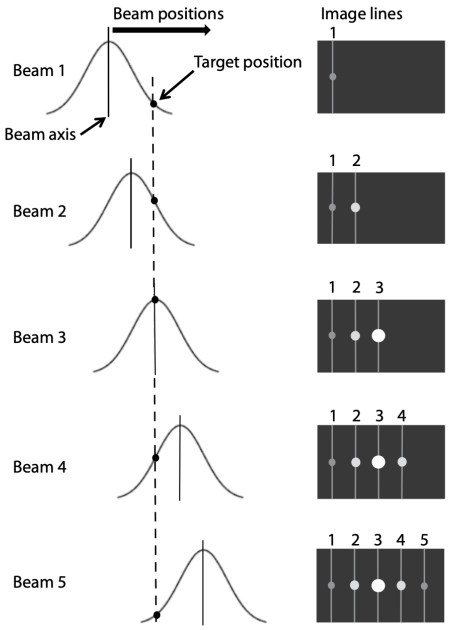



**Fig. 1.13.** The lateral response profile describes the lateral intensity profile generated when imaging a point scatterer. For a scanline image, i.e., an image where multiple beams are transmitted from a set of sub-apertures with constant relative lateral offsets, this profile results from the width of the transmitted beam being non-zero when encountering the scatterer, and therefore lateral response being registered. The Figure above visually describes the origin of this profile with beam position shifted over the scatterer for the individual scanlines transmitted and discrete lines in image space depicting the resulting intensity as circles whose brightness represents the relative amplitude of the response [64].

and results in a loss of focus and overall more noise in the ultrasound image. Other artifacts include shadowing and amplification. Shadowing occurs when an interface's attenuation or acoustic impedance, often bone, is so large that no signal from below the interface returns to the transducer, leading to a large black region or shadow below the interface. Conversely, amplification occurs when a region in the image, often a water-filled organ, has lower attenuation than expected, and signals are brighter behind the region. Many of these artifacts are today thought to be characteristic of the ultrasound modality but represent the current state of the art. By discovering ways to create an adaptive ultrasound imaging modality that is more aware of the medium it is currently interrogating, many of these artifacts could be reduced, thereby improving the overall image quality possible with ultrasound imaging.

The following passage will discuss the metrics with which ultrasound image quality is measured.

### 1.4.4 Image Quality Metrics

A selection of quality metrics is used to quantitatively evaluate the imaging properties between transducers, targets, and scans. These metrics measure resolution, or the smallest distance at which objects can be differentiated and contrast, the range of intensities between light and dark objects.

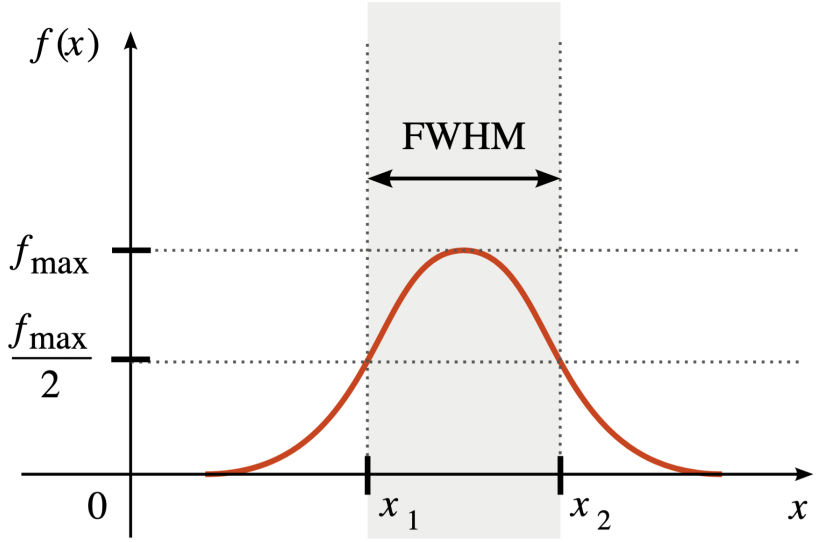



**Fig. 1.14.** For a given distribution, full width half maximum (FWHM) describes the difference between two independent variables whose value is half the maximum of the distribution. This metric is often used in signal processing to describe when two beams can be considered separate. See Figure 1.15 for a practical application. Image sourced from Wikipedia under the GNU Free Documentation License, Version 1.2, https://commons.wikimedia.org/wiki/File:FWHM.svg

Resolution can be measured both axially, along the beam propagation path, and laterally, orthogonal to the propagation path in the imaging plane. Laterally resolution is commonly measured with the full-width half maximum of the PSF. This metric defines the lateral width at which the intensity of the PSF, generated from the pulse-echo response of a single scatterer or point target, drops -6dB or $\approx 0.5$, leading to the name.

Contrast refers to the differentiation between intensities of two neighboring regions [64]. Given two homogeneous regions $B_L$ of lesion intensities and $B_B$ of background intensities. The contrast ratio (CR) can be defined as the difference of the mean of both regions normalized by the mean of the background region and can be written as:

$$CR = \frac{\mu(B_L) - \mu(B_B)}{\mu(B_B)}$$

This definition works well for homogeneous cases, but in cases where noise makes viewing a lesion more complex, an additional metric can be helpful.

Contrast-to-noise ratio (CNR) is closely related to CR, but the denominator takes into account the standard deviation of the background region [64]. CNR is formally defined as:

$$CNR = \frac{\mu(B_L) - \mu(B_B)}{\sigma(B_B)}$$

Quantitative metrics in ultrasound images are still actively investigated to find more robust metrics that are less dependent on medium and imaging parameters [142].

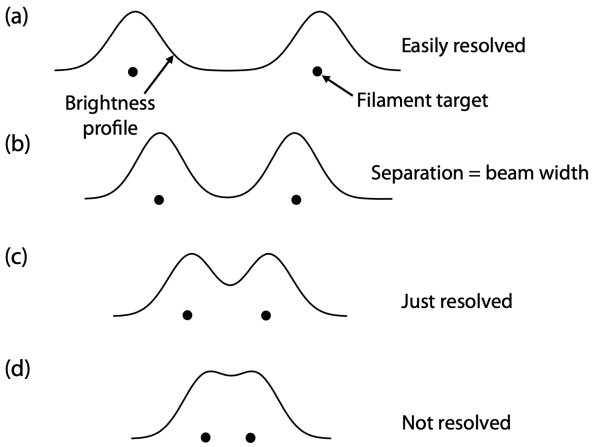



**Fig. 1.15.** Subplots (a)-(d) show a progressive transition from resolved scatterers to unresolved scatters. Subplot (a) displays two points with a large lateral offset and above the curve of their lateral response signal. The two peaks of the two response signals are separated, meaning that the two points can be resolved in the resulting ultrasound image. Progressively, moving through the subplots, the point targets move closer together, and the resulting lateral response profile becomes less and less resolved. In subplot (c), the two response profiles have begun to overlap, but critically, the full width half maximum (FWHM) of the two profiles is still separable. Ultimately, subplot (d) shows that the two points can no longer be resolved since the two peaks can no longer be differentiated from one another [64].

## 1.5 Clinical Application

Medical ultrasound has many clinical applications and can image a plethora of anatomies safely and inexpensively. Breast ultrasound imaging is performed in conjunction with mammography to investigate indeterminate lesions and can be highly effective in identifying breast lesions [157]. These exams are performed by a trained radiologist with a linear probe. Distinguishing features that one might look for in breast ultrasound to classify malignant lesions are rough edges, position in the breast, texture (e.g., speckle within the boundaries of the lesion), and blood perfusion [136]. Furthermore, thyroid screenings with ultrasound can also identify and locate thyroid nodules, which can cause imbalances in the endocrine system and potentially be malignant. Again here, a visual inspection with a linear probe can already give reasonable first indications lesion type [190]. Other examples of clinical evaluation include cardiac prostate ultrasound, which uses niche probe layouts to image their target anatomy.

In general, medical ultrasound applications are plentiful, and transducer shape and element layout can be adjusted to the task at hand. Furthermore, advanced imaging techniques exist, such as Doppler imaging which can measure object velocity within an ultrasound frame, and shear wave elastography, which can indicate tissue elasticity in-vivo, further expanding the range of applications for diagnosis and treatment of medical ultrasound. Still, the radio frequency signals of medical ultrasound encode more information than the intensities, which are currently visualized in an ultrasound image. The following passages of this work will discuss the emergence of deep learning in the field of computer vision and current and future applications of deep learning on unprocessed or weakly processed ultrasound signals.

# 2 Deep Learning in Natural Images and Ultrasound

**Contents**



Deep learning describes the body of algorithms that mimic the structure of the mammalian cerebral cortex process to model complex and non-linear function [19]. Deep learning is a component of the broader field of machine learning methods that work towards artificial intelligence. Deep learning is based on deep feed-forward networks, which can also be referred to as feed-forward neural networks or multi-layer perceptions (MLPs). These modeled are constructed to behave as a universal approximator [31, 60, 63, 99] and approximate an arbitrary function $f$ where $\mathbf{y} = f(\mathbf{x})$ given an input $\mathbf{x}$ and label $\mathbf{y}$. The model learns the parameters $\theta$ such that $\mathbf{y} = f(\mathbf{x}; \theta)$. These approximators can be linked, creating a stack of functions. Recent studies have shown the potential of deep networks, with more parameters, for better performance on a variety of tasks from computer vision [181], to speech recognition [35], natural language processing [193], bioinformatics [111], machine translation [178] and more. This Chapter will discuss the composition of these universal approximators, the methods used to train them, and some basic applications in computer vision. We will then discuss how these methods can be applied to improve ultrasound imaging.

## 2.1 Deep Learning in Natural Images

As mentioned, Deep Learning describes a connectionist system by which a long chain of neural network layers maps a function given input data $x$ to a corresponding output $y$. The parameters $\theta$ of each layer commonly consist of a linear layer including a set of weights $w$ and a bias $b$. The layer can then be written as

$$f(\mathbf{x}; \mathbf{w}, b) = \mathbf{x}^\top \mathbf{w} + b.$$

In order to be able to generalize to non-linear functions, a source of non-linearity is required called an activation function. Historically this activation function has been a sigmoid function [117]. Still, more recently, it has been shown that a rectified linear activation unit (Relu) is less computationally expensive and can therefore accelerate the training process [47, 74, 116]. Networks consist of a collection of layers, and therefore, for our example, we will add a second layer $f^{(2)}$ with weights $\mathbf{W}$ and bias $\mathbf{c}$ to our first linear layer, now designated $f^{(1)}$.

When we add these components to our original linear model above, we get our two-layer feed-forward network.

$$f(\mathbf{x}; \mathbf{W}, \mathbf{c}, \mathbf{w}, b) = f^{(2)} max\{0, f^{(1)}(x)\} = \mathbf{W}^\top max\{0, \mathbf{w}^\top \mathbf{x} + b\} + c.$$

When trained, $f(\mathbf{x}; \mathbf{W}, \mathbf{c}, \mathbf{w}, b)$, will be able to approximately map a function between the input $x$ and output $y$ on which it was trained. For simplicity of formulation, the parameters of the neural network $f$ will be summarized as $\theta$.

The cost function, also often referred to as criterion, loss function, error function, or objective function, describes a functional formulation of a property that one would like to enforce. This enforcement functions by minimizing the cost function of the output of a given neural network with respect to the neural network parameters $\theta$. The cost functions in deep learning are often well known from classical machine learning applications. There are many examples of cost functions, such as optimizing distribution constraints with a maximum likelihood cost function. For discrete-valued outputs, a cross-entropy cost function can be effective. For continuous values, a mean-absolute-error or mean squared error cost function can be advantageous, though they might also lead to poor training results [49]. For the imaging task of segmentation, often the Sørensen–Dice coefficient (DSC) is employed, which can be defined as:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

Training a neural network is similar to other gradient descent optimization problems. The only difference is that the non-linearity of the neural network makes the system highly non-convex. Neural networks can still be trained in an iterative gradient-based fashion via **stochastic gradient** decent, which can be applied to loss functions with no convergence guarantee. **Stochastic gradient** decent replaces traditional gradient descent with an estimation of traditional gradient descent. The idea behind stochastic gradient descent is that the system is iteratively minimizing an objective function on batches of data given a cost function instead of the entirety of the data at once. To use stochastic gradient descent for optimization, the objective function must fulfill certain smoothness properties, such as being differentiable or sub-differentiable. This optimization problem can be written with $\mathcal{J}_i(\theta)$ defining the loss function of the $i$th batch of the data as:

$$\theta := \theta - \eta \nabla \mathcal{J}(\theta) = \theta - \frac{\eta}{n} \sum_{i=0}^{n} \nabla \mathcal{J}_i(\theta).$$

In this formulation of stochastic gradient descent, $\eta$ represents the step size of the optimization problem. The step size is often referred to as the learning rate in machine learning. The stochastic gradient descent approach has the advantage that the entire data set does not have to be loaded into memory at one time. Furthermore, it is based on the assumption that a batch of the training data is large enough to be statistically representative of the data set as a whole. By training iteratively on batches, one reduces the computational burden by trading off a lower convergence rate for faster iterations [13].

Now that we have briefly discussed network construction, optimization methods for training, and loss definition, it is important to discuss the gradient generation methods required for training. When a neural network is passed an input $\mathbf{x}$ and through a series of matrix multiplication and non-linearity steps generates an output $\mathbf{y}$, this step is called *forward-propagation*. After forward propagation, the scalar cost value $\mathcal{J}(\theta)$ can be generated. Given this scalar value, the *backpropagation* algorithm [145] can be applied to allow the cost information to flow backward through the network to produce the gradient.

Backpropagation models the above formulated neural network as a computation graph to reduce the computational complexity of gradient calculation for deep neural networks, thereby making the problem computationally tractable. Though the analytical expression of a gradient may be trivial, the numerical computations of gradients can prove expensive and, at times, intractable. To apply backpropagation to the shallow neural network formulated above, each variable in the network is assigned a node in a graph. Operations define the connections between nodes and thus how each node is related. The gradient of the neural network can be broken down into sub-expressions, defining the operations required to traverse from one node to the next. When calculating the entire graph's gradient, one must merely calculate the gradients of individual components and combine them via the chain rule.

The entire training process using backpropagation can be summarized as follows: Given a ground truth sample $\hat{y}$ and a training sample $x$, the first forward pass is performed to get a network estimation $\hat{y}$. The forward pass refers to the iterative multiplication of the sample $x$ with the weights $W^{(i)}$ and biases $b^{(i)}$ of the layers $i \in \{0 \dots l\}$ to step by step calculate the activations $a^{(i)}$. Once a network estimation $\hat{y}$ is generated, the estimation is compared to the ground truth $y$ via the loss function $\mathcal{J}$. Now we are ready to compute the gradients of the network. First, we compute the gradient of the output layer by differentiating the loss function with respect to the estimation $\hat{y}$. We then traverse the graph backward, differentiating every operation with respect to its output and multiplying the current gradient with the gradient of the parent node. We repeat this process until every node has been differentiated and we have reached the original input $x$. After the gradients have been computed, an algorithm such as gradient descent or other relevant optimization algorithm is used to update the weights and biases of the network.

Training MLP neural networks as described above is a simplified example. As the number of parameters of a neural network grows, the ability of the network to map to more and more complex problems is improved. Adding more layers to a neural network, and thereby improving the learning capability of the network, birthed the field of deep learning (DL), or the training of deep neural networks (DNN) [90]. Nevertheless, this development also increased training complexity and inferring with these neural networks. Furthermore, in computer vision applications, many of the learned weights and filters must be repeated for every section of the image data. To reduce the complexity of neural networks and reduce the number of repeated filters learned, the idea of convolutional neural networks (CNN) was proposed [89, 91]. CNNs can learn a set of spatial filters that can be convolved over the activation of the layer input. The use of CNNs reduces the complexity of neural networks by convolving the learned filter over the entire layer input, therefore allowing deeper networks, accelerated training, and better generalization.

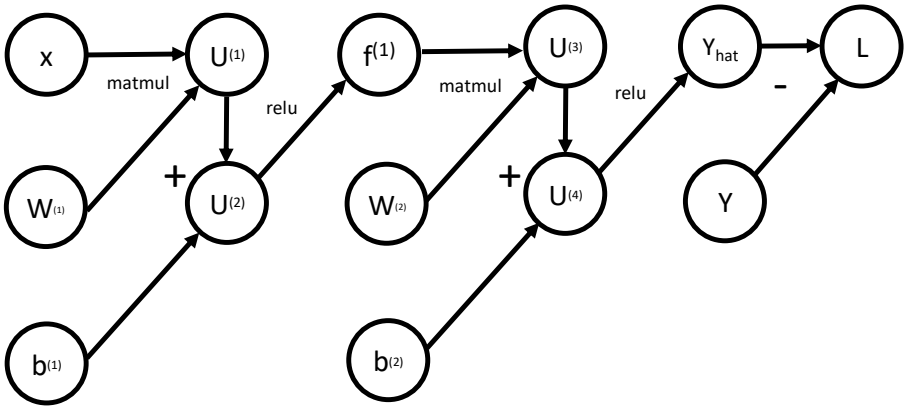



**Fig. 2.1.** A simple multi-layer perceptron (MLP) is represented as a graph. Each node of the graph represents a data state, while each node represents a data operation. The intermediate states are represesnted by $U_i$. The output of the MLP is compared with a ground truth $\hat{y}$ to calculate the loss value, which can subsequently be back-propagated.

With this high-level understanding of the composition of neural networks, we can briefly look at how these algorithms have been applied in the field of computer vision and medical imaging before exploring their application in the field of medical ultrasound imaging.

Applications of CNNs and DNNs are diverse and plentiful. In the field of computer vision, CNNs have been used for semantic image segmentation [98], simultaneous localization and mapping (SLAM) [113, 159], depth estimation [184], classification [87], object detection [196], face recognition [34], trajectory planning and estimation [124] and human pose estimation [169]. The application of deep learning in these fields has led to significant performance gains in the technology and its massive adoption in computer-aided medicine. Applications in medicine range from 3D volumetric segmentation of MRI and CT [110, 143], disease classification and diagnosis [195], medical robustness analysis [125], CT and MRI reconstruction [56, 81, 191], outcome and disease prediction [20], medical data augmentation [3, 126], surgical phase recognition and workflow analysis [32], treatment planning [39], medical robotic ultrasound navigation [58].

## 2.2 Deep Learning in Ultrasound

There has been much interest in applying neural networks to ultrasound images as with other medical imaging modalities. Due to the proportion of the interfaces being imaged to the wavelength of ultrasound images the many artifacts, there are many potential applications for deep learning in ultrasound imaging that have been explored. Applications include pre-processing, ultrasound reconstruction (beamforming) for improved image quality [100, 102, 115, 129, 148, 179], similar to super-resolution in natural images [192], accelerated reconstruction of ultrasound signals [84, 101, 155], multi-focus imaging with generative adversarial neural networks [54], as well as ultrasound image speckle filtering [69, 70], cluster and reverberation filtering [16], and quantitative ultrasound applications [40]. Here, we will briefly describe the approaches to combine the fields of ultrasound beamforming and deep learning.

### 2.2.1 Deep Learning for ultrasound beamforming

Ultrasound imaging has unique requirements due to the medical application. For example, the transducer is separate from the computational hardware and moves relative to the processing hardware. Unlike with natural images where a CCD/CMOS is statically positioned physically close to the processing hardware, ultrasound transducers are required to be portable to accommodate the shape and location of the human bodies they image. Furthermore, the computational power in ultrasound devices is often not generalized and dynamic but a configured set of parameters in hardware, e.g., a field-programmable gate array (FPGA). This led to significant latency and limited compute power for image reconstruction and beamforming. Furthermore, due to the significant amplitude discrepancies in ultrasound imaging between highly reflective interfaces and diffuse scattering media, higher precision of either 32 or 64 bits is required in ultrasound imaging than the industry standard of 12 to 14 in natural imaging. This leads to high bandwidth requirements in medical devices, which add to cost and complexity. Nevertheless, physicians have an expectation and medical requirement of real-time, meaning 30 frames per second.

Delay and sum beamformers are the current industry standard for real-time imaging applications. At receive time, a set of delays are applied to the channel data to focus on a point in space, and the signals over the aperture are summed. During this process, the time of flight is assumed constant. When these beamforming assumptions about the required delays break down, discrepancies in the wave's travel time through the tissue can cause imaging artifacts and loss of resolution.

**Ultrasound Signal Processing**

Several data-driven approaches have been presented to combat these issues with the standard beamforming pipeline, which learns methods that filter ultrasound signals or learn a set of delays based on the received ultrasound data. [83, 154, 155] proposed training and fully convolutional encoder-decoder network that maps pre-delayed channel data to beamformed output for improved image quality. Other methods proposed advanced filtering of channel data that can be classically beamformed in a subsequent step [16]. The suppression of off-axis scattering was also explored by learning a spectral filtering method with a multi-layer perceptron [100]. Hyun et al. proposed the use of a fully convolutional network to learn to filter speckle in beamformed ultrasound images, thereby improving resolution [69]. The network was passed 17 sub-aperture RF signals beamformed as input and returned a speckle-reduced B-mode image. Other applications used generative adversarial networks [53] to synthesize multiple-focus images from single focus images. In total, many approaches have been presented by which data-driven neural networks are trained to filter or reconstruct ultrasound data, trained on data alone. But some think that the knowledge of the reconstruction process should not be left out of the equation but rather integrated into the deep learning process by adding models in the loop.

**Model-Based Deep Learning Approaches**

In order to constrain the solution space of the deep learning problem for image enhancement one can borrow methods from adaptive beamforming techniques such as Capon of Minimum

variance beamforming [22, 149, 160, 180]. In Capon beamforming, a set of weights are optimized for such that the variance of the signal across the aperture is minimized. This can be done by solving the set of equation

$$\hat{\mathbf{w}} = \arg\min \mathbf{w}^H \mathbf{R}_x \mathbf{w} \tag{2.1}$$

$$\text{s.t. } \mathbf{w}^H \mathbf{a} = 1. \tag{2.2}$$

Here $\mathbf{R}_x$ is the covariance matrix calculated over the receiving elements while $\mathbf{a}$ denotes the steering vector of the transmission. For perfectly delayed signals, $\mathbf{a}$ is a normalized unit vector.

Solving Equation 2.2 requires the inversion of $\mathbf{R}_x$, whereby the computational complexity is cubic relative to the number of elements [15]. To address this computational complexity with deep learning, [103] proposed learning the inversion of the covariance with four fully connected layers. The input for the network was predelayed channel data for a given pixel in image space, and the output a corresponding set of channel apodization weights $\mathbf{w}$. The network has access to a large amount of training data since pixels are processed independently and lead to a 400x speedup [103]. The proposed method resulted in reduced clutter and improved resolution in the resulting images [177].

Though much work has been done in the field, deep learning has still only experienced limited acceptance in the field of medical ultrasound research. While neural network architectures and methods have boomed in the field of computer vision, in which they were developed, in ultrasound, methods have until now been adapted to the modality, rather than reinvented from the ground up for the task of acoustic imaging. Many prefer traditional reconstruction and beamforming methods, despite their shortcomings, to a black-box approach using a deep neural network. What is still required is a method that merges the statistical priors and understanding a deep neural network can provide, with the understanding that today's models have of the physics of ultrasound wave propagation, scattering, attenuation, etc.

# Part II

Ultrasound Simulations

# Ultrasound Simulations

# 3

## Contents



## 3.1 Problem Statement

In ultrasound research and development, as in other physical sciences, simulation has become a helpful tool for technology development [76, 94, 146, 163, 171, 186, 197]. Computer-based simulation tools reduce the amount of physical experimental setup required to evaluate a hypothesis and therefore allow faster iteration and new technologies. These simulation algorithms can range in complexity and realism, dependent on the application at hand.

Specifically, in the case of ultrasound, simulations allow the user to generate a set of signals for a given transducer medium pair without the physical requirement of either the transducer or the medium. Until now, ultrasound simulations allowed researchers and developers to generate radio frequency data, with which they could develop and evaluate image reconstruction and beamforming algorithms [28, 76, 146, 171].

One popular such simulation suite is Field-II, which simulates the spatial impulse response given a medium (phantom) and transducer pair [76]. In Field-II, the medium is parameterized by scatterer density and echogenicity values of the medium. Until now, Field-II was an excellent tool to generate radio frequency signals for a given medium transducer pair to develop beamforming image reconstruction methods.

Today, the field of quantitative ultrasound methods is of growing interest [121]. Researchers are eager to investigate the relationships between the physical properties of the interrogated medium and the resulting radio frequency signal. For such applications, researchers require a simulation framework that can model the physical tissue properties and not only their echogenicity. In this case, the Field-II simulation framework does not model this complexity, and more complex simulation frameworks need to be examined.

There is a wide variety of numerical simulation software for computation fluid simulation. When selecting one for a given application, one has to evaluate the trade-offs of numerical accuracy and computational complexity. For a given application, it is more of an art than a

science when selecting simulation software. Factors that influence the selection of an ultrasound simulation software include but are not limited to the size of the computational domain, the number of frequencies of interest, the medium properties, the boundary conditions of the simulation, and the sensitivity of the application to numerical simulation artifacts [171]. Specifically for the investigation of quantitative ultrasound methods with medical ultrasound, we are interested in the solution of the wave equation that can simulate heterogeneous media and provide a time-domain solution of the pressure over time. A full computational fluid dynamics solver was formulated by Gianmarco Pinton and published under the name Full-wave [131, 132, 133]. For an acceptable level of accuracy, such approaches are computationally expensive and require as many as 10 points per wavelength for satisfactory simulation results. For three dimensional simulations, this requirement can lead to highly accurate simulations, but at the cost of long-run times [171, 172].

The k-Wave simulation software offers a solution to this problem by solving a system of coupled first-order partial differential equations with a global *k*-space pseudo-spectral method [11, 12, 108]. Since the basis functions are sinusoidal, only two grid points are required per wavelength rather than ten for the full computational fluid dynamics method [133, 171]. This simplification reduces the computational complexity and makes k-Wave an attractive alternative to a full computational fluid dynamics suite.

The k-Wave framework simulates a spatial, temporal wave equation. It can base the resulting simulated signals on physical inputs such as sound speed, density, non-linearity, and attenuation intensity maps. This foundation in the physical principles of wave propagation allows researchers to use k-Wave to bridge the gap between the physical world of quantitative properties and the resulting radio frequency signals that modern ultrasound transducers generate.

## 3.2 k-Wave

The k-Wave framework or k-Wave toolbox is a collection of tools and functions for simulating time-domain acoustic wave propagation in 1D, 2D, and 3D. Originally developed at University College London (UCL) by Bradley Treeby and Ben Cox and was first released in 2009, the software is flexible and can simulate linear and non-linear wave propagation through a heterogeneous medium. To date, the k-Wave toolbox has been written with a MATLAB user interface and both MATLAB and C++ computational loops. The C++ computational loops supported but distributed and accelerated high-performance computing (HPC) tasks and were added to the project by Jiri Jaros of Brno University of Technology [171]. In its most complete form (c.f. Equations pressure-density conservation) the k-Wave package is able to solve the Westervelt Equation [165, 187].

### 3.2.1 Practical application of k-Wave

To better understand the contributions to the modeling of realistic biological tissue in Chapter 4, the practical parameterization of a k-Wave simulation will be briefly discussed. This section should give the reader an overview of the required inputs for a successful k-Wave simulation.

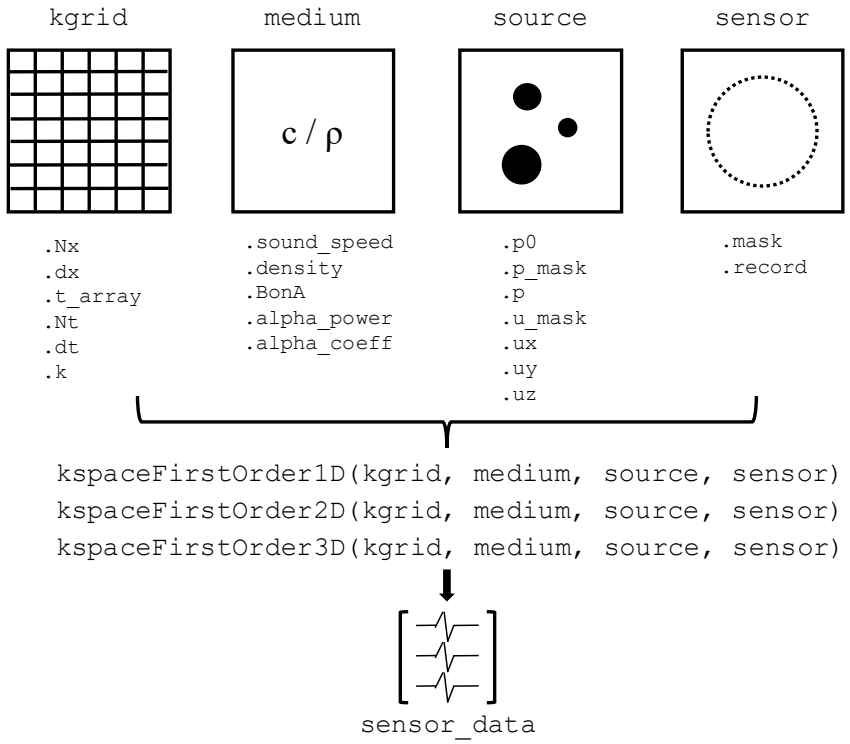



**Fig. 3.1.** Visual class diagram of k-Wave simulations. The simulation method requires four-object variables to run, namely a kgrid, a medium, a source layout, and a sensor layout. The objects and the properties of the objects are listed in the diagram above.

k-Wave simulations are parameterized with four general input objects: a computational grid or **kgrid**, a medium, a source, and a sensor. The **kgrid** object defines the spatial and temporal discretization of the simulation. The medium defines the spatial distribution of the physical parameters of the medium. These include sound speed, density, non-linearity ("B over A"), attenuation power, and the attenuation coefficient. Further, k-Wave requires the definition of so-called "sources" and "sensors". Sources and sensors define the mask of grid points, on which source terms are added, and the resulting temporal pressure signals are recorded.

### 3.2.2 Numerical Model and Governing Equations

There are many states in the medium one can model to model acoustic wave propagation, such as pressure, density, temperature, particle velocity, etc. In the quiescent[1] and isotropic[2] case, the wave equation defines the relation of these terms in a second-order partial differential equation:

$$\nabla^2 p - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} = 0 \tag{3.1}$$

where $p$ is the pressure field, $c_0$ defines the wave propagation speed through the medium. The wave equation can be decomposed into a set of coupled first-order partial differential equations that dictate the conservation of momentum and mass as well as the pressure-density ratio [130]. It is from these equations that Equation 3.1 can be derived.

[1] Quiescent describes a voxel with constant boundary conditions and no net flow in or out of the voxel [48]
[2] Isotropic illustrates the fact that the wave propagation properties are independent of propagation direction

$$\begin{aligned}
\frac{\partial \mathbf{u}}{\partial t} &= -\frac{1}{\rho_0}\nabla p, && \text{(momentum conservation)} \\
\frac{\partial \rho}{\partial t} &= -\rho_0 \nabla \mathbf{u}, && \text{(mass conservation)} \\
p &= c_0^2 \rho && \text{(pressure-density relation)}
\end{aligned} \tag{3.2}$$

We first model the second-order wave equation as the set of first-order differential equations, which allows the addition of the source terms of mass and force into the modeled systems and particle velocity, which can be used in subsequent calculations to model medium temperature change due to wave propagation.

### Attenuation

In k-Wave, the medium is further modeled as an attenuating fluid, meaning part of the energy of the wave propagation is converted to heat over the distance traveled. This is modeled by the frequency power law:

$$\alpha = \alpha_0 \omega^{\eta}, \tag{3.3}$$

where $\eta$ is the power-law exponent, $\alpha_0$ is the power law pre-factor $[Np(rad/s)^{-ym^{-1}}]$, and $\omega$ is the angular frequency [6, 18]. When the attenuation term is added to the system of first-order differential Equations pressure-density relation, and they take the following form:

$$\begin{aligned}
\frac{\partial \mathbf{u}}{\partial t} &= -\frac{1}{\rho_0}\nabla p, && \text{(momentum conservation)} \\
\frac{\partial \rho}{\partial t} &= -\rho_0 \nabla \cdot \mathbf{u} - \mathbf{u} \cdot \nabla \rho_0, && \text{(mass conservation)} \\
p &= c_0^2(\rho + \mathbf{d} \cdot \nabla \rho_0 - L\rho). && \text{(pressure-density conservation)}
\end{aligned} \tag{3.4}$$

Here, $\mathbf{d}$ is the acoustic particle offset or displacement. The operator $L$ in the pressure-density term of Equations pressure-density conservation is a linear integro-differential operator which is added to account for dispersion and absorption following the frequency power law following [183], which states that to obey causality, acoustic absorption must be physically accompanied by dispersion. The $L$ operator can be written out as:

$$L = \tau \frac{\partial}{\partial t}(-\nabla^2)^{\frac{y-1}{2}-1} + \eta(-\nabla^2)^{\frac{y-1}{2}+1}, \tag{3.5}$$

where the absorption and dispersion proportionality terms can be written as

$$\tau = -2\alpha_0 c_o^{y-1} \text{ and, } \eta = 2\alpha_0 c_0^y \tan(\pi y/2) \tag{3.6}$$

Further, to keep the validity of the mass conservation term under the observation of the adjusted pressure-density term, the $\nabla \rho_0$ terms cancel each other out.

### Non-Linearity

As described in Chapter 1, non-linear wave-propagation occurs when, for sufficiently large amplitude waves, then the pressure differential between wave peaks and wave troughs is large

enough to modify the medium sound speed in the peaks and troughs of the wave, respectively. This results in an acceleration of the wave peaks and a retardation of the wave troughs and, therefore, a non-linear distortion of the waveform [45]. To account for some but not all of these effects, k-Wave adds $\frac{B}{A}$ to the discretized model. The term $\frac{B}{A}$, spoken "B over A," characterized the first two terms of the Virial Expansion [45] and parameterized the influence of amplitude-dependent non-linear effects on sound speed. The non-linearity term can be added to the model equations such that:

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} &= -\frac{1}{\rho_0}\nabla p, && \text{(momentum conservation)} \\ \frac{\partial \rho}{\partial t} &= -(2\rho + \rho_0)\nabla \cdot \mathbf{u} - \mathbf{u} \cdot \nabla \rho_0, && \text{(mass conservation)} \\ p &= c_0^2(\rho + \mathbf{d} \cdot \nabla \rho_0 + \frac{B}{2A}\frac{\rho^2}{\rho_0} - L\rho). && \text{(pressure-density conservation)} \end{aligned} \tag{3.7}$$

In this case, a mass conservation equation is augmented with an additional term to account for convective non-linearity in which the particle velocity influences the wave-velocity [57]. The additional term in the pressure-density equation accounts for the material non-linearity. Together, all the Equations pressure-density conservation can be simplified to a generalized form of the Westervelt Equation [165, 187].

### Source Terms

The generalized Westervelt equation formally defines the behavior of a wave in a medium, but without an initial condition within the medium, no wave would propagate. To generate waves in the heterogeneous medium, source terms must be added to the equation formulations. These source terms can be added as either mass or force sources.

The main difference between force and mass source terms is the directivity of the sound fields they generate. The formulation of their addition will be discussed in the following sections.

#### Force Source Terms

Force source terms result from a force being applied in the direction defined by the force vector. In this way, force sources are directional. The resulting field is therefore generated a dipole field. Examples of force sources are pistons oscillating or transducers. The addition of a force source term assumes the addition of an acceleration $[ms^{-2}]$. In k-Wave, force sources are added as velocity terms in the momentum-conservation equation.

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} &= -\frac{1}{\rho_0}\nabla p + \mathbf{S}_F, && \text{(momentum conservation)} \\ \frac{\partial \rho}{\partial t} &= -(2\rho + \rho_0)\nabla \cdot \mathbf{u} - \mathbf{u} \cdot \nabla \rho_0, && \text{(mass conservation)} \\ p &= c_0^2(\rho + \mathbf{d} \cdot \nabla \rho_0 + \frac{B}{2A}\frac{\rho^2}{\rho_0} - L\rho). && \text{(pressure-density conservation)} \end{aligned} \tag{3.8}$$

**Mass Source Terms**

Mass source terms generate a monopole field, i.e., concentric pressure waves spreading from the source position. An example of a mass source term is oscillating bodies, e.g., a buoy floating on the surface of a body of water. Mass source terms are added to the mass conservation equation and have the unit [kg $m^{-3}s^{-1}$]. Within k-Wave mass, source terms are applied as a temporally changing pressure field over time.

$$\begin{aligned}
\frac{\partial \mathbf{u}}{\partial t} &= -\frac{1}{\rho_0}\nabla p, && \text{(momentum conservation)} \\
\frac{\partial \rho}{\partial t} &= -(2\rho + \rho_0)\nabla \cdot \mathbf{u} - \mathbf{u} \cdot \nabla \rho_0 + \mathbf{S}_M, && \text{(mass conservation)} \\
p &= c_0^2(\rho + \mathbf{d} \cdot \nabla \rho_0 + \frac{B}{2A}\frac{\rho^2}{\rho_0} - L\rho). && \text{(pressure-density conservation)}
\end{aligned} \tag{3.9}$$

### 3.2.3 Pseudo-Spectral Numerical Solver

To simulate ultrasonic signals accurately based on acoustic source terms and the known distribution of medium properties, the correct specification of a simulation regime is required. We are interested in the time-domain solution of the wave equation for broadband acoustic waves in heterogeneous media. The decision on a simulation regime is based upon numerical stability, computational complexity, memory complexity, and scaling and parallelization behavior. Temporal computational fluid dynamic simulations can be extremely accurate but require at least 10 points per wavelength for an accurate and stable simulation [133]. This leads to poor scaling behaviors despite simulation accuracy.

Pseudo-spectral methods for solving systems of differential equations can be advantageous due to the numerical simplification of some operations, and subsequent acceleration of the simulation and have been proposed for the scattering wave equation [11, 12, 108, 162]. Pseudo-spectral methods simplify some operators by transforming the formulation of the system of equations into the spectral domain. The spectral domain is often considered the frequency domain in signal processing but can also be evaluated in the spatial-spectral domain of k-space. This transformation does require that the system domain be formulated as a periodic domain, meaning that the boundary conditions at opposing sides of the domain are coupled. To simulate real-world three-dimensional domains with pseudo-spectral methods, a dampening layer is placed between the boundary condition to attenuate signals that pass through the periodic boundary. In the context of our discussion in this section, this layer is referred to as the perfect-matching-layer (PML)[3].

In this case, the spectral transformation facilitates the calculation of spatial derivatives with a temporal propagator in the spatial frequency of the k-space domain. Psuedo-spectral methods and their application in heterogeneuous ultrasound simulation have been well discussed

[3]Unfortunately, the collision of two fields of research is problematic here. While in this work, we refer to the PML as to the dampening boundary layer around our spatial domain to formulate it periodically, in the world of ultrasound transducers that we are simulating, PML refers to the perfect-matching-layer between the acoustic source elements and the tissue.

and evaluated [11, 12, 14, 30, 41, 42, 43, 51, 52, 108, 162, 167, 173]. Through in-depth discussion of these numerical methods goes beyond the scope of this work, the spectral formulations of the linear case equations in Section 3.2.2 which were used in the scope of this work will be listed for the sake of completeness.

The linear case of the wave equation formulated in Equation pressure-density relation the mass and momentum equations with added sources can be written using a k-space pseudo-spectral method. With the notation of $n$ and $n+1$ denoting the current and next time-step, respectively, this method can be written as:

$$\frac{\partial}{\partial \epsilon} p^n = \mathcal{F}^{-1}\left\{ ik_\xi \kappa e^{i\kappa_\xi \Delta\xi/2} \mathcal{F}\{p^n\}\right\}, \tag{3.10a}$$

$$u_\xi^{n+\frac{1}{2}} = u_\xi^{n-\frac{1}{2}} - \frac{\Delta t}{\rho_0}\frac{\partial}{\partial \xi} p^n + \Delta t \mathrm{S}_{\mathrm{F}_\xi}^{\mathrm{n}}, \tag{3.10b}$$

$$\frac{\partial}{\partial \xi} u_\xi^{n+\frac{1}{2}} = \mathcal{F}^{-1}\left\{ ik_\xi \kappa e^{-ik_\xi \Delta\xi/2} \mathcal{F}\left\{u_\xi^{n+\frac{1}{2}}\right\}\right\}, \tag{3.10c}$$

$$\rho_\xi^{n+1} = \rho_\xi^n - \Delta t_{\rho_0}\frac{\partial}{\partial \xi} u_\xi^{n+\frac{1}{2}} + \Delta t \mathrm{S}_{\mathrm{M}_\xi}^{n+\frac{1}{2}}. \tag{3.10d}$$

Equations 3.10a and 3.10c depict the spatial gradient calculation derived from the Fourier collocation spectral method, and 3.10b and 3.10d depict the k-space corrected first-order accurate forward difference update step. These steps are performed in an N dimensional space of $\mathcal{R}^N$ where $\xi \subset \{x, y, z\}$, i.e. the set of spatial directions. The $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent the forward and inverse spatial Fourier transform, $i$ denotes the imaginary unit, $k_\xi$ denotes the wavenumber in direction $\xi$. The grid spacing is written as $\Delta\xi$ for direction $\xi$, and the time-step is written $\Delta t$. The k-space operator $\kappa$ is defined as:

$$\kappa = \mathrm{sinc}(c_{ref}) k \Delta t / 2,$$

where $c_{ref}$ is the reference sound speed. The discrete wave numbers $k_\xi$ are defined as

$$k_\xi = \begin{cases} \left[-\frac{\mathrm{N}_\xi}{2}, -\frac{\mathrm{N}_\xi}{2}+1, \ldots, \frac{\mathrm{N}_\xi}{2}-1\right]\frac{2\pi}{\Delta\xi \mathrm{N}_\xi} & \text{if } \mathrm{N}_\xi \text{ is even} \\ \left[-\frac{(\mathrm{N}_\xi-1)}{2}, -\frac{(\mathrm{N}_\xi-1)}{2}+1, \ldots, \frac{(\mathrm{N}_\xi-1)}{2}\right]\frac{2\pi}{\Delta\xi \mathrm{N}_\xi} & \text{if } \mathrm{N}_\xi \text{ is odd} \end{cases}.$$

for the number of grid points $\mathrm{N}_\xi$ in the $\xi$ direction. The term in Equations 3.10a and 3.10c of $e^{\pm ik_\xi \Delta\xi/2}$ is a spatial shift operator to offset the gradient result calculations by half a grid point and is denoted by $n \pm \frac{1}{2}$, which allows to evaluate the particle velocity components on a staggered grid as illustrated in Figure 3.2. This staggered grid can increase accuracy and stability when computing odd-order derivatives [42].

The pressure-density relation is understood to be the ambient density defined at the staggered points and is given by:

$$p^{n+1} = c_0^2(\rho^{n+1} - L_d),$$

given a total acoustic density of

$$\rho^{n+1} = \sum_\xi \rho_\xi^{n+1}.$$

The source terms in Equations 3.10b and 3.10d denote the input forces per mass unit and temporal rate of mass input per unit volume. To input acoustic pressure and velocity, the required values are scaled from pressure and velocity inputs. $\mathrm{S}_{\mathrm{F}_\xi}$ and $\mathrm{S}_{\mathrm{M}_x}$ are calculated as:

$$\mathrm{S}_{\mathrm{F}_\xi} = u_\xi \frac{2c_0}{\Delta\xi} \text{ for } \xi \in \{x, y, z\} \text{ and} \tag{3.11a}$$

$$\mathrm{S}_{\mathrm{M}_\xi} = \frac{p_\xi}{c_0^2 N} \frac{2c_0}{\Delta\xi} \text{ for } \xi \in \{x, y, z\} \tag{3.11b}$$

In the non-linear case, the convective non-linearity term is added to Equation 3.10d and can be written as:

$$\rho_\xi^{n+1} = \frac{\rho_\xi^n - \Delta t \rho_0 \frac{\partial}{\partial \xi} u_\xi^{n+\frac{1}{2}}}{1 + 2\Delta t \frac{\partial}{\partial \xi} u_\xi^{n+\frac{1}{2}}} + \frac{\Delta t \mathrm{S}_{\mathrm{M}_\xi}^{n+\frac{1}{2}}}{1 + 2\Delta t \frac{\partial}{\partial \xi} u_\xi^{n+\frac{1}{2}}}. \tag{3.12}$$

Due to the temporal gradient in the mass conversion term (c.f. Equation pressure-density conservation) being solved using an implicit finite difference scheme, a non-linear correction term is applied to the mass source term. Since this effect is small on the source term, it is not applied to the source term.

The corresponding pressure-density relation can be written to include a non-linearity term as:

$$p^{n+1} = c_0^2 \left( \rho_{n+1} + \frac{\mathrm{B}}{2\mathrm{A}} \frac{1}{\rho_0} - \mathrm{L}_d \right), \tag{3.13}$$

again with a total acoustic density of $\rho^{n+1} = \sum_\xi \rho_\xi^{n+1}$.

In order to add frequency dependent absorption into the numerical model, as was introduced in Section 1.2.9, a fractional Laplacian is added to the model to account for the frequency dependency[26, 170], which can be computed efficiently in Fourier spectral methods when compared to temporal fractional derivatives [23, 25, 82, 96, 161, 188]. The spatial Fourier transform of the negative fractional Laplacian can be written as [26, 134]:

$$\mathcal{F}\{(-\nabla^2)^a \rho\} = k^{2a} \mathcal{F}\{\rho\}, \tag{3.14}$$

which leads to the discritized form of the absorption term for the power law to written as [**12**]:

$$\mathrm{L_d} = \tau \mathcal{F}^{-1} \left\{ k^{y-2} \mathcal{F} \left\{ \frac{\partial \rho^n}{\partial t} \right\} \right\} + \eta \mathcal{F}^{-1} \left\{ k^{y-1} \mathcal{F} \left\{ \rho^{n+1} \right\} \right\}. \tag{3.15}$$

For computational efficiency, the temporal derivative of the acoustic density can be replace with a linearized mass conservation equation of $\frac{\partial \rho}{\partial t} = -\rho \nabla \cdot \mathbf{u}$ and allows us to write Equation 3.16 as:

$$\mathrm{L_d} = \tau \mathcal{F}^{-1} \left\{ k^{y-2} \mathcal{F} \left\{ \frac{\partial \rho^n}{\partial t} \right\} \right\} + \eta \mathcal{F}^{-1} \left\{ k^{y-1} \mathcal{F} \left\{ \rho^{n+1} \right\} \right\}. \tag{3.16}$$

Lastly, we will briefly discuss the numerical PML for the periodic computational domain of the pseudo-spectral computational methods as was used in the scope of this work. As mentioned, to stop waves leaving the periodic simulation domain on one side to reappear on the opposing side, a layer of highly attenuating material is added to the periodic boundary. This layer

attenuates any departing wave and reduces its appearance on the opposite side. There are two critical requirements for the PML. This layer must a.) provide enough absorption is attenuation sufficiently, and b.) not reflect any waves back into the medium. Using the split-field formulation of perfect matching layer from Berenger [8, 9, 61], we can write the first-order coupled equations with the perfect matching layer terms as:

$$\begin{aligned}
\frac{\partial \mathbf{u}_\xi}{\partial t} &= -\frac{1}{\rho_0}\frac{\partial p}{\partial \xi} - \alpha_\xi u_\xi, && \text{(momentum conservation)} \\
\frac{\partial \rho_\xi}{\partial t} &= -\rho_0 \frac{\partial \mathbf{u}_\xi}{\partial \xi} - \alpha_\xi \rho_\xi, && \text{(mass conservation)} \\
p &= c_0^2 \sum_\xi \rho_\xi, && \text{(pressure-density conservation)}
\end{aligned} \tag{3.17}$$

such that $\alpha = \{\alpha_x, \alpha_y = 0, \alpha_z = 0\}$ is the anisotropic absorption. In accordance, with [162, 194] one can transform the momentum and mass conservation equations into the form of:

$$\frac{\partial}{\partial t}(e^{\alpha_\xi t} u_\xi) = -e^{\alpha_\xi t}\frac{1}{\rho_0}\frac{\partial p}{\partial \xi}, \qquad \frac{\partial}{\partial t}(e^{\alpha_\xi t}\rho_\xi) = -\rho_0 e^{\alpha_\xi t}\frac{\partial u_\xi}{\partial \xi}.$$

With a first-order forward differences discretisation scheme, Equations 3.10a and 3.10b can be brought into the form used in k-Wave simulations and written as:

$$u_\xi^{n+\frac{1}{2}} = e^{\alpha_\xi \Delta t/2}\left(e^{-\alpha_\xi \Delta t/2} u_\xi^{n-\frac{1}{2}} - \frac{\Delta t}{\rho_0}\frac{\partial}{\partial \xi} p^n\right), p_\xi^{n+1} = e^{-\alpha_\xi \Delta t/2}\left(e^{-\alpha_\xi \Delta t/2}\rho_\xi^n - \Delta t \rho_0 \frac{\partial}{\partial \xi} u_\xi^{n+\frac{1}{2}}\right).$$

Lastly, to reduce reflection on the boundary, the attenuation rate is be annealed with:

$$\alpha_\xi = \alpha_{\text{max}}\left(\frac{\xi - \xi_0}{\xi_{\text{max}} - \xi_0}\right)^m,$$

with $\xi_0$ representing the start of the perfect matching layer and $\xi_{\text{max}}$ the end. In the simulations in this work, the setting of $m = 4$ is used following [162].

Now that we have covered the numerical formulation of the simulation environment employed in this work, we will discuss the generation and systematic parameterization of simulation mediums that created realistic time-domain radio frequency ultrasound signals.

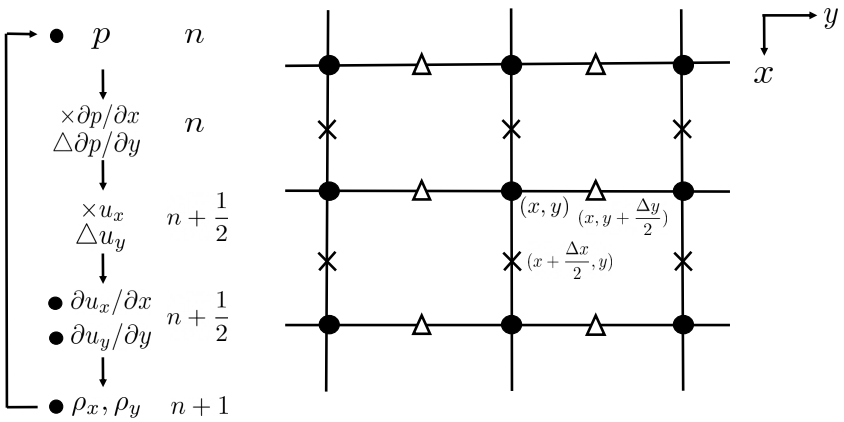



**Fig. 3.2.** Schematic diagram of the computation steps to simulate ultrasound signals using a pseudo-spectral coupled first-order approach. $\frac{\partial p}{\partial x}$, $\frac{\partial p}{\partial y}$ and $u_x, u_y$ are positioned at staggered grid points laterally and vertically and denoted by triangles and crosses. All other variables are calculated at the dots on the grid. The time step at which each variable is solved for is denoted by $n$, $n + \frac{1}{2}$ and $n + 1$.

# 4 Simulation Medium Contributions

**Contents**



## 4.1 Simulated Medium

The following Chapter describes the methodology used to create realistic in-silico simulations for generalizable training of neural networks with simulated radio frequency data. To this end, we will cover the generation and parameterization of a realistic in-silico breast phantom, the simulation process using the k-Wave suite, the proposed data processing, and augmentation steps for training a deep neural network architecture and structure of the DNN. The use of the k-Wave simulation suite allows the creation of a data set with a paired sound speed and density medium sample. The k-Wave toolbox, though powerful, still requires careful parameterization to achieve realistic ultrasound simulations that are comparable to in-vivo measurements. Our method for ultrasound simulations ensures that the resulting simulations have realistic physical properties and numerically optimized and stable execution. The physical properties we ensure include a fully formed speckle pattern, realistic echogenic intensity variations between tissue types, speckle response from sound speed and density variations, proper accounting of realistic non-linearity and attenuation properties, and a correct and accurate transmit steering. Numerically, the suitability of the simulations is ensured via a proper automated accounting for a suitable Courant–Friedrichs–Lewy (CFL) condition number for a given simulation and an optimized runtime, thanks to the low prime factors of the grid size when the k-Wave perfect matching layer (PML) is accounted for.

The ultrasound simulations developed in this work are generated to model human breast tissue and are comprised of three basic elements; a scatterer distribution field, a tissue variation model for the background of breast images, and a random spatial distribution of anatomical features in the image composed of skin, lesions, and background. The scatter distribution field dictates the location and intensity of random scatterers in a medium. The skin layer models the tissue properties of skin and the anatomical depth skin normal displays. The simulated lesions are made to model echogenic and anechoic lesions. Lastly, the background layer is generated to model the echogenicity and geometry of the subtle variations breast tissue can

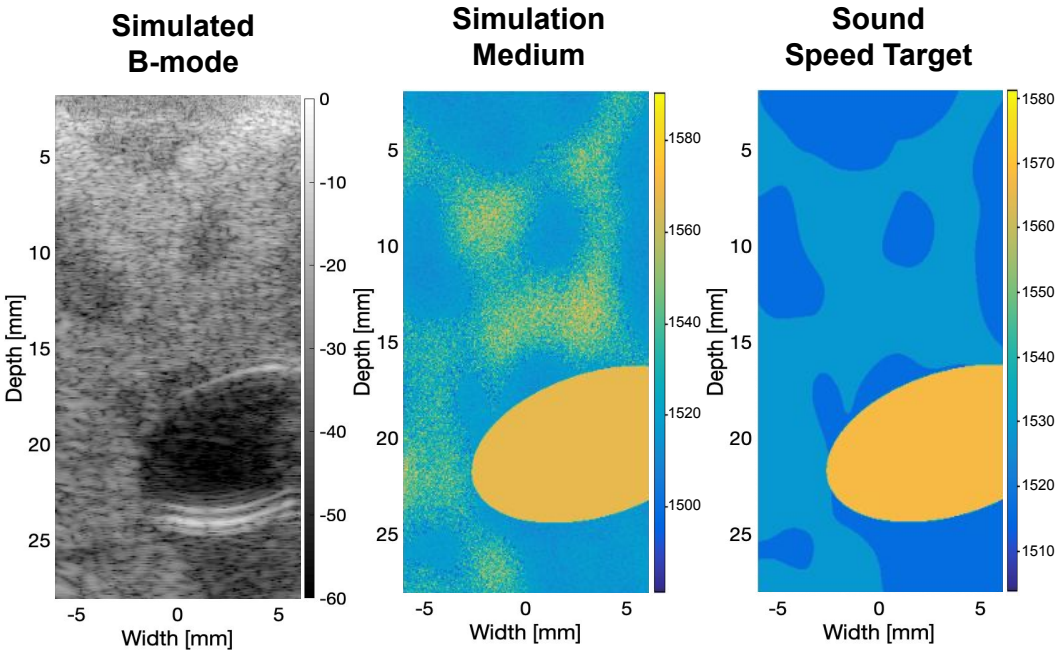



**Fig. 4.1.** (Left) Simulated ultrasound B-mode image with background approximating glandular breast tissue and an anechoic cyst [156]. (Middle) Sound speed of the simulated medium. (Right) Sound speed target used for model optimization with region-average sound speed values. Note that two sound speed values are used for the background, and a single sound speed value is used for the cyst.

display. The combination and random parameterization of these elements generate a large heterogeneous dataset.

### 4.1.1 Medium Domain

The in-silico phantom domain is defined on a Cartesian grid of points $p_i \in X \times Y \times Z$, where

$$X = \{0, x, 2x, ..., x_d\}, \tag{4.1}$$

$$Y = \{0, y, 2y, ..., y_d\}, \text{ and} \tag{4.2}$$

$$Z = \{0, z, 2z, ..., z_d\}. \tag{4.3}$$

Here, $x$, $y$, and $z$ are the spatial resolution of the grid in the respective directions and $x_d$, $y_d$ and $z_d$ are the respective grid dimensions. A mapping is assumed from

$$p_i \mapsto (x_p, y_p, z_p)$$

from the set of points in the Cartesian grid to the spatial dimensions the Cartesian grid resides within.

### 4.1.2 Scatterer Distribution

The simulation scatterer density

$$\rho_s = min(\frac{n_s}{\lambda^3} \cdot x \cdot y \cdot z, 1) = min(\frac{f_t^3 \cdot n_s}{c_0^3} \cdot x \cdot y \cdot z, 1), \tag{4.4}$$

where $n_s$ is the number of scatterers in an imaging resolution voxel of size $\lambda^3$ and $\lambda$ is the wavelength $\lambda = \frac{c_0}{f_t}$. Here $c_0$ is the assumed imaging sound speed of the simulation, and $f_t$ is the transmit frequency of the transducer. For every point $p_i$ a random independent and identically distributed sample (i.i.d.)

$$u_i = U \cdot B \text{ where } U \sim \mathcal{U}_{[-0.5,0.5]} \text{ and } B \sim \mathcal{B}(1, \rho_s),$$

is sampled to create a spatial white noise distribution that defines the position and relative amplitude of scatterers in the domain. The Bernoulli distribution $\mathcal{B}(1, \rho_s)$ models the scatterer density $\rho_s$.

### 4.1.3 Tissue Classes

To realistically model breast anatomy, skin, breast gland, breast cysts, and breast lesions resembling fibroadenomas and glandular tissue are simulated [156].

First, the breast gland tissue is generated to model the variation in background echogenicity found in breast ultrasound images and serves as the background of our simulated medium.

A 2D Gaussian filter $g$ of the size

$$(x_f, y_f) \in \{(j,k) : j \in [|X|], k \in [|Y|], j \text{ and } k \text{ even } \}$$

is defined as

$$g(u,v) = \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}}$$

where

$$u \in [-\frac{x_f}{2}, \frac{x_f}{2}] \text{ and } v \in [\frac{-y_f}{2}, \frac{y_f}{2}].$$

The filter $g$ is then convolved with a 2D random field of size $F = [0, x_d + x_f] \times [0, y_d + y_f]$ where $F_i \sim \mathcal{U}_{[0,1]}$. The resulting value field is then normalized ($\mu = 0$) and re-scaled to the interval $[-0.5, 0.5]$. This is the basis of the subsequent breast gland model and is later scaled with the mean sound speed and augmented with the scaled scatterer map.

Next, both cysts and lesions in the dataset are modeled by elliptical inclusions but are differentiated by the fact that cysts are anechoic while lesions can have either positive or negative echogenicity. Cysts and lesions are the most common abnormalities in breast tissue and can appear in women of any age and change during the menstrual cycle [156]. The cyst/lesion mask is defined as an ellipse in space projected on to the aforementioned Cartesian

grid and parameterized by the position of its center $p_c \in [X \times Y]$ and the lengths of its radii $r_{i=\{1,2\}} \in \mathbb{R} \quad \forall \quad r_i < min(x_d, y_d), \quad$ and a random orientation angle $\theta \in [0, \pi]$.

$$E(m_i = (x_p, y_p)) = \frac{((x_p - x_c) \cdot \cos(\theta) + (y_p - y_c) \cdot \sin(\theta))^2}{r_1} + \frac{((x_p - x_c) \cdot \sin(\theta) - (y_p - y_c) \cdot \cos(\theta))^2}{r_2} \tag{4.5}$$

Finally, skin tissue is simulated for varying thickness within anatomical norms of 0.7 to 3 mm [66].

### 4.1.4 Property Assignment

For a given class region, the background sound speed is scaled to the desired mean value. Additionally, the scatterer field intensity is also scaled to achieve the desired echogenicity and added to the background. The ranges of the sound speed and contrast for each given class can be seen in Table 4.1 and were chosen following [5]. The density map is scaled proportionally to the sound speed map by a factor of $\alpha_\rho$ and attenuation and non-linearity are set to a constant value.

In total, six combinations of the above-mentioned tissue classes are formed, namely, cyst with skin, lesion with skin, skin, background (breast gland), lesion, and cyst. Lastly, the in-silico phantom sound speed map is averaged by region (two sound speeds for background, one for cyst/lesion and one for skin) to form a coarse target sound speed map, as shown in Figure 4.1, suitable for training our deep model. These in silico phantoms are then utilized in k-Wave to generate simulated RF channel signals from pulse-echo ultrasound.

**Tab. 4.1.** Mean sound speed range and scatter contrast per class used for our breast ultrasound dataset simulation.

| Property | Mean Sound Speed Range | Scatter Contrast |
|---|---|---|
| Cyst [5] | [1500, 1620] | - |
| Lesion [5] | [1488, 1512] | ± 10-30 dB |
| Skin | [1540, 1670] | 10 dB |
| Breast Gland [5] | [1480, 1528] | 12 dB |

## 4.2 Results and Discussion

Based on the simulation method described in the previous section, one can simulate ultrasound signals and their resulting signals with a full-wave simulation. The described method allows for a realistic final image and signals with minimal simulation artifacts. These are two important requirements for deep learning on ultrasound data when data transfer from the simulated domain to the real world is desired. Figure 4.2 displays the resulting simulated B-modes and

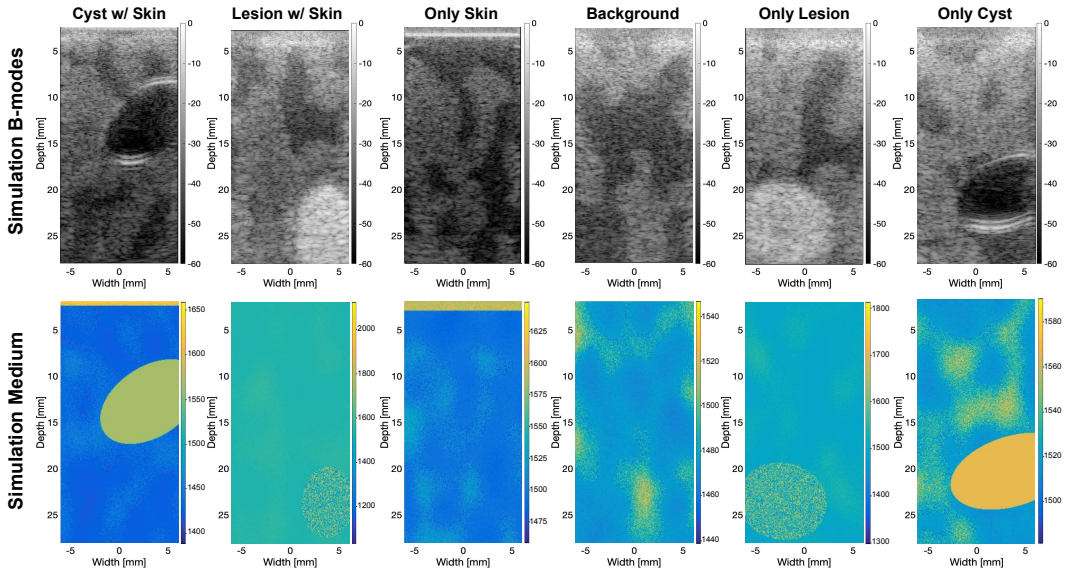



**Fig. 4.2.** Simulated B-mode images from each of our six classes along with their simulation medium. Our simulations produce realistic B-modes showing contours of cysts, lesions, skin and background variations.

their respective input media. A fully-fledged speckle distribution can be seen in the B-modes, along with accurate contrast and texture distributions. The simulation pipeline is able to automatically generate randomly sampled tissue types and accurate B-modes. Each column in Figure 4.2 is one of the six classes which can be generated.

In general, the requirements of ultrasound simulation frameworks evolve as the applications of the simulated data change. Until now, simulations are often used to test algorithms in common and simple cases, such as homogeneous speckle, circular anechoic cysts, and arrays of point scatterers, often under perfect sound speed conditions. Rarely, are the requirements of simulations such that the simulation must match the property distribution of in-vivo images. With the application and training of deep learning algorithms on simulated data, simulations must be more realistic for the trained networks to be evaluated with real transducers and in-vivo data.

# Part III

Tissue Sound Speed Estimation

# 5 Sound Speed Estimation with Deep Learning

## Contents



## 5.1 Introduction

Sound speed estimation in ultrasound imaging has the potential to radically change how ultrasound imaging is used in medical practice but remains a challenging problem. In general, ultrasound continues to evolve from its foundation of B-mode image reconstruction. Many new modalities have been developed with ultrasound imaging that has helped better quantify the tissue of interest in the past decades. These include Doppler imaging [166], strain imagine [174], shear wave elastography [4, 164], functional ultrasound [104]. Furthermore, for breast lesion classification, methods have been proposed to classify malignant breast lesions directly on RF time-series data [176]. These augmenting modalities aid clinicians

in their diagnosis by providing further quantitative measures used in downstream statistical diagnostic models.

Doppler imaging allows physicians to measure the relative velocity of a medium relative to the transducer via the well-understood Doppler effect [166]. In pulse-Doppler imagine, the frequency shift in the transmitted pulse is proportional to the relative velocity of the interrogated medium in motion. This information is presented to clinicians as a color-coded field on top of their grey-scale B-mode image.

Shear wave elastography generates information about the biomechanical properties of the tissue by generating transverse shear wave pulses that propagate through the tissue of interest at imaging time[46, 55, 119, 164]. The tissue displacement magnitude is subsequently measured via speckle tracking algorithms from which the shear modulus is subsequently derived.

Lastly, functional ultrasound [104] uses high-speed plane-wave imaging to extract a blood signal $s_B$ from a series of compounded plane-wave images by applying a high-pass filter to the image intensity values over time. The spectrum of $s_B$ gives information on relative blood velocity over time. This new technology has been shown to work in proof of concept studies on mice.

Though these quantitative methods have been significant developments in the field of ultrasound, one important quantity can still not be estimated in a clinical setting: sound speed. Sound speed or speed of sound is the physical property that directly influences ultrasound resolution. Sound speed defines the heterogeneous propagation speed of a wave through a medium and can vary from point to point. Since the variation of sound speed is currently not modeled in image reconstruction techniques, it can lead to signal degradation in images and all other quantitative ultrasound methods. Therefore a robust method for sound speed estimation would be an important milestone in developing ultrasound imaging technologies and a landmark in the clinical efficacy of handheld ultrasound imaging.

### 5.1.1 Problem Statement

Currently, the most advanced sound speed estimation methods rely on ultrasound tomography (UT), which requires specialized hardware consisting of either a linear transducer that is mechanically rotated around a region of interest or a convex transducer into which the tissue of interest can be placed. For total tomographic sound speed reconstruction, a complete angular sampling from $[-\pi, \pi]$ is required [80] and tomographic reconstructions require that signals transmitted from a transducer be received by a sensor positioned opposite. For this reason, tomographic sound speed imaging is challenged by the large attenuation in bony tissue, beyond which signal intensity is greatly decreased[135, 147]. Pulse-echo, sound speed estimation, is one way to overcome these challenges and generate a sound speed distribution without the burden of specialized hardware.

Pulse-echo sound speed estimation describes the process by which an ultrasonic transducer is positioned over a region of interest, and one or more ultrasonic waves interrogate the medium

to generate a sound speed estimate. Using the pulse-echo response from the tissue, a model is created by which a sound speed distribution within the tissue can be derived. In general, these models can be separated into two sub-categories: (1) physical model methods, which use the RF or IQ channel measurement from the transducer array as input and solve a linear system of physical (2) machine learning models and data-driven approaches, which are trained on a series of paired data (either RF or IQ) to predict a sound speed distribution in the tissue being imaged.

## 5.2 Current Methods in Sound Speed Estimation

### 5.2.1 Physical Model-based Approaches

Anderson and Trahey [2], building on initial studies [141] and inspired by applications in seismology [151] proposed a novel, intuitive method by which a global average sound speed between the transmitting interface and a focal point could be derived from features extracted from the channel data of a focused transmit. This work was a first approach at global average sound speed estimation with pulse-echo ultrasound. In this method, a focused pulse was fired at the desired tissue depth. Then a pulse-echo response of the focused wavefront was received. Assuming a perfect orthogonal one-wave geometric wave profile, the delay profile of the returning signal can be modeled by fitting a best-fit curve to the signal. One key assumption of the Anderson Trahey method is the orthogonal propagation direction of the focused transmit wave relative to the transducer face. Assuming a given Cartesian coordinate space defined relative to a linear transducer aperture with the $x$, $y$, and $z$ axis corresponding to the lateral, elevational, and axial dimensions, the geometric one-wave delay profile $t(x)$ of the returning signal from a wave focused at $(x_t, y_t, z_t)$ in a medium with a sound speed $c$ can be described as:

$$t(x) = \frac{\sqrt{(x - x_t)^2 + y_2^2 + z_t^2}}{c}. \tag{5.1}$$

Anderson et. al. show that given the known positions of the transducer elements $x$, this formulation can be simplified an one can solve for the unknowns $c$, $x_t$, $y_t$ and $z_t$ by squaring each side of the Equation 5.1 and rearranging the terms in to a polynomial to the form $f(x) = p_1 x^2 + p_2 x + p_3$ where

$$p_1 = \frac{1}{c^2}, \quad p_2 = -\frac{2x}{c^2}, \quad p_3 = \frac{x_t^2 + x_t^2 + y_t^2}{c^2}. \tag{5.2}$$

They show that one can then fit a second-order polynomial fit of $t^2(x)$ in the least-squares sense in order to determine $p_1, p_2$ and $p_3$ given the assumption that $y_t \ll z_t$. The resulting formulations can be written as:

$$c = \frac{1}{\sqrt{p_1}}, \quad x_t = -\frac{c^2 p_2}{2}, \quad z_t \approx \sqrt{c^2 p_3 - x_t^2}. \tag{5.3}$$

This method was evaluated experimentally on four different fluids and two separate speckle-generating phantoms. The phantom's sound speeds ranged from 1136 m/s to 1547 m/s and were all made with a commercial ultrasound device. The resulting mean relative error of the experiments was less than $\pm 0.4\%$. The method was initially only evaluated on homogeneous phantom models, where the scattering medium was placed directly in the focal region. The technique was shown to struggle when presented with strong inhomogeneities [2]. In general, the assumption of a known reflector position was shown not to hold in highly heterogeneous media, and the modeling of tissue non-uniformity was left for future work.

Building on the work of Anderson and Trahey, Jakovljevic et al. [73] proposed a physically-inspired model that enabled the estimation of a local sound speed map along an orthogonal wave propagation path from a sequential series of global average sound speed measurements at discretized depths. To do this, Jakovlijevic et al. measured a grid of global average sound speeds with a grid spacing of half-a-wavelength and 130 x 64 total measurements (axial x lateral).

The work models the total average sound speed to a given point as the discretized sum of the local average sound speeds between two discreet points or:

$$c_{avg} = \frac{1}{N} \sum_{i=1}^{N} c_i, \tag{5.4}$$

i.e., the average sound speed to a given discrete point equals the arithmetic mean of the local sound speeds along the travel path. For all discrete points in the grid and all elements in the transducer, this relationship can be formalized into a linear system of equations where:

$$c_{\text{avg}} = A c_{\text{local}} + \varepsilon_{\text{meas}}. \tag{5.5}$$

Here, $c_{\text{avg}}$ contains the a single measurement of average sound speed in every row, and the model matrix $A$ models the relationship of $c_{\text{local}}$ to $c_{\text{avg}}$ along with $\varepsilon_{\text{meas}}$ models the system error.

Though in some special cases, $A$ can be lower triangular and can be solved directly, often, given a set of samples, Equation 5.5 was solved via gradient descent with a quadratic least-squares regularizing term. The method was validated with perfect arrival times calculated with the eikonal equation and evaluated in both full-wave simulations and phantom trials. In the full-wave simulations, the resulting method had a bias between 3 and 4.3 m/s and a standard deviation of 0.3 m/s for experiments performed on 1520 m/s 1540 m/s, and 1570 m/s phantoms, respectively. The model was successfully shown to work when the arrival times and therefore the $c_avg$ are known. Experiments were performed on full-wave simulation discussed in Chapter 3, and a two-layered phantom which showed promising results in the absence of noise.

Nevertheless, significant hurdles still exist to applying channel data methods in the medical workflow with inhomogeneous media. First, the model assumes straight line wave propagation and does not account for the lateral propagation of the beam or the effects of refraction or diffraction. This assumption introduces error into the measurements' already ill-posed optimization problem. Furthermore and critically, the method assumes a homogeneous media

to fit the returning wavefront to a second-order polynomial and thereby extract travel times, which would be a challenge to circumvent in inhomogeneous media, for example, in-vivo measurements of the abdomen or breast.

Building on the work of Jakovlijevic et al. and working to address the shortcomings of previous approaches, Ali and Dahl proposed the IMPACT method that was better able to estimate sound speed in volumes with axial inhomogeneities. This was made possible by modeling ray paths from every point in the reconstruction domain to every transducer element and not just the axial wave propagation through the medium [1]. The tomographic reconstruction model employed by IMPACT still neglected refraction used a straight ray approximation to model the travel time. The travel time $t_i$ to a focal point $(x_f, z_f)$, given an element position $(x_i, 0)$ was modeled as:

$$\tau_i(x_f, z_f) = \int_0^{D_i(x_f,z_f)} \frac{dr}{c\left(x_i + \frac{r(x_f - x_i)}{D_i(x_f,z_f)}, \frac{rz_f}{D_i(x_f,z_f)}\right)} \tag{5.6}$$

where the path length is defined as $D_i(x_f, z_f) = \sqrt{(x_f - x_i)^2 + z_f^2}$. By integrating the propagation lines indexed for every transducer element, the authors create a linear system of equations:

$$\overrightarrow{t}_{obs} = \boldsymbol{H}\overrightarrow{s} \tag{5.7}$$

where $\overrightarrow{s}$ is the vector of pixels in the sound speed reconstruction, $\boldsymbol{H}$ is the system matrix describing the relationship between pixels and interrogation rays for a given element. The global average sound speed for each pixel in the domain $\overrightarrow{t_{obs}}$ defines the left-hand side. To calculate the global average sound speeds for the reconstruction grid, a grid search was employed from 1400 m/s to 1700 m/s to find the average sound speed that maximized the speckle coherence factor at that point; a method proposed by [105]. The coherence factor (CF) is defined as:

$$CF = \frac{|\sum_{k=1}^{N} s[k]|^2}{N\sum_{k=1}^{N} |s[k]|^2}, \tag{5.8}$$

where $s[k] \in \mathbb{C}$ is a complex sample from the element $n \in [0, N]$ for a given imaging point. Mallart and Fink initially showed in [105] that this term in a homogeneous medium has a maximum value of $\frac{2}{3}$ given a cylindrical focus and is decreased given uncompensated aberration inhomogeneities. In the IMPACT method from [1], the CF images are spatially smoothed to obtain a speckle-averaged coherence. With this speckle-smoothed global average sound-speed map, a local sound speed distribution is approximated with a tomographic model defined in Equation 5.6. Quantitative and qualitative results displayed in the paper loaoked promising. The authors even showed the potential improvement in image reconstruction by calculating delay times via the Eikonal Equation [17] and displayed improved contrast and half-maximum width. Still, the method remained sensitive to significant lateral variations in the sound speed distribution.

Sanabria et al. [147] solved the inverse problem directly in the spatial domain with a novel anisotropically-weighted total-variation method for regularization. In this work, the forward problem was constructed as a differential time-of-flight measurement based on apparent displacement and a given wave along a ray-like propagation path. For the forward process, a given element-to-element propagation path $p \in P$ is given, where $P$ is defined as the set of all

propagation paths. The time of flight (TOF) delay $t_p$ is defined on a discretized grid [x,z] as:

$$t_p = \int_C \sigma dl \approx \sum_{c=1}^{C} l_{p,c}\sigma_c. \tag{5.9}$$

Here, $l_p, c$ is defined as the path length of path $p$ from the transmitter TX to cell $c$ in the grid and back to the receiver RX. The slowness $\sigma_c$ is equal to the inverse of the discretized sound speed on the grid or $\sigma_c = v_c^{-1}$. This model does not hold when the structure of the interrogated medium is not accurately known, and therefore one can write a formulation of the relative time measurements $\tau_m$ along different paths as:

$$\tau_m = \sum_{p=1}^{P} \omega_{m,p} t_p \tag{5.10}$$

where for each measurement $m = 1, .., M, \tau_m$ describes the relative time delays between paths $t_p$. The weights $\omega_{m,p}$ are defined as ternary weights i.e. $\omega_{m,p} = -1, 0, 1$. In this work, they successfully displayed high-resolution sound-speed reconstructions for both accurate time of flight measurements and measurements corrupted by noise.

More recently, Stähli et al. [158] extended the CUTE method [72] by solving for sound speed maps with a system of spatially distributed phase shift measurements taken between pairs of transmit and receive angles (Tx and Rx) set around a common mid-angle. In the proposed method, a complex radio frequency (crf) image is reconstructed for given transmit angles $\theta_{tx}$ and receive angles $\theta_{rx}$ with a common mid-angle. This reconstruction choice is made due to the fact that signals with a common midpoint are well correlated regardless of the underlying scatter distribution upon which the reflection is based. This approach, therefore, circumvents the potentially anisotropic nature of the reflected response from a given isochronous point.

The method further considered the erroneous position of the echos in the reconstruction. Due to the aberration delay, the exact position of a reflector's true location is unknown at the time of reconstruction. To reconcile this issue, this work assumes that all reflected signals are received with an angle pair $(\theta|\psi)$ and derives a function from characterizing the offset difference between the reconstruction position and the true spatial position in the medium. They define this offset $d$ as:

$$d = \frac{\hat{c}[\tau_{tx} + \tau_{rx}(n, g)]}{2\cos[\frac{1}{2}(\theta_n - \psi_{n,g}]}, \tag{5.11}$$

where $\tau_{tx}$ and $\tau_{rx}(n, g)$ are the transmit and receive delays to a point, and $n$ and $g$ indexing the transmit parameters and common mid angle respectively and $c$ representing the assumed beamforming sound speed. To paper goes on to connect the positional offset to a measured phase shift and updates the total model equation to:

$$\Delta\Theta(r', n, n'g) \simeq 2\pi f_0 \left\{ \frac{\tau_{tx}(n') + \tau_{rx}(n', g)}{cos\left[\frac{1}{2}(\phi_{n'} - \psi_{n',g}\right]} - \frac{\tau_{tx}(n) + \tau_{rx}(n, g)}{cos\left[\frac{1}{2}(\phi_n - \psi_{n,g})\right]} \right\}. \tag{5.12}$$

The equation is parameterized by the aberration correction sample location $r'$, corrected by the offset $d$, the transmit elements $n$ and $n'$ that reach the position $r'$ simultaneously. The variables $\phi$ and $\psi$ define the transmit and receive angles, respectively.

This novel approach created accurate sound speed maps in a series of phantoms and displayed marked improvements over previously published baselines. Furthermore, an in-vivo reconstruction on a liver model exhibited realistic sound speed estimates. Nevertheless, the authors state that since CUTE is based on phase tracking, it can be sensitive to motion artifacts in in-vivo imaging but proved robust when a breath-hold procedure was applied.

In general, physics-based methods that depend on iterative solvers like conjugate gradient methods to solve the respective system of equations are challenging to apply in real-time pulse-echo imaging. Also, though straight line approximations of wave propagation are useful in modeling, they reduce the realism of the system model. Therefore, it is essential to investigate methods that do not need solver convergence at the time of measurement and generalize beyond ray approximations.

### 5.2.2 Deep Learning-based Approaches

The second class of single-sided sound speed inversion methods is built upon the recent advances in computer vision with the advent of neural networks as universal estimators. Specifically, convolutional neural networks have been shown to provide high-quality estimations for tasks such as the segmentation and classification of natural images. Convolutional neural networks also benefit from the advantage of constant scaling for arbitrary image size [89]. Lastly, network building blocks such as skip connection, batch-normalization [71] as well as training regimes such as the ADAM optimizer [85] have improved the speed and efficiency with which neural networks can be trained. These improvements lend themselves also to the application of convolutional neural networks as universal approximators for the estimation of sound speed images.

Feigen et al. [40] was one of the first to propose single-sided sound speed estimation with a deep neural network. In his work, the network architecture employed was based on VGG [153], which was initially proposed in 2014 to investigate the effect of depth on convolutional neural network performance and evaluated on the ImageNet data set. The network was trained to map the raw channel data from three plane wave transmits to a simulated sound speed distribution when applied to sound speed estimation.

The transmit protocol for the channel data consisted of three sub-apertures of 64 elements on a 128 element linear transducer. The first aperture consisted of elements 1-64, the second elements 32-96, and the third elements 64-128. The second transmission had a steering angle of 0 degrees while the others lateral transmissions were steered with symmetrical angles $\pm x^{\circ}$ with and angle $x^{\circ}$ that was "chosen to best cover the full domain" [40]. The transmission depth was set to 4 cm.

The data dimensions passed to the model were $[transmit, elements, time - samples]$. In the paper, three different variations of encoders were investigated. The first network, dubbed "start network," took the data above and encoded the transmit dimension as channels to the network. This stacked data had no spatial correlation between the channel data since a separate steering angle was used for each transmission. The second network architecture variant was called "middle," concatenated three encodings in the channel dimension after the

first eight convolutional filters and passed them through one decoder. The last network variant was called “end” and encoded the first eight convolutions and decoded five convolutions separately. After the fifth decoding convolutions, the data representations were concatenated in the channel dimension and passed through one final convolutions filter to conclude this network variation.

The training data was generated by simulating ultrasonic interrogations of media containing ellipses of varying ultrasonic properties and randomly positioned in a domain with the k-Wave software package [171]. The soft-tissue simulation medium was generated on a four-dimensional grid. The four dimensions consisted of three spatial dimensions and one property dimension for the various input medium properties of the simulation, including sound speed, density, non-linearity, and attenuation. The medium grid was set to have a homogeneous background with constant density of $0.9\frac{g}{cm^3}$.

In this domain, between one and five ellipses were randomly placed in the sound speed dimension. The sound speed of the randomly selected ellipses was randomly sampled from a range of $1300m/s$ to $1800m/s$. Random scatterers were distributed independently of the aforementioned background or ellipses within the density dimension. The scatterer density was stated to vary randomly between $-3\%$ and $6\%$ of the mean density. The scatter distribution density, i.e., the spatial density with which the scatterers were placed, was stated to be two reflectors per wavelength squared, despite as was stated in Section 4.1.4, it is a rule of thumb to have at least ten reflectors per wavelength squared resolution voxel to have a fully developed speckle (FDS) in the resulting ultrasound image. A fixed attenuation was was set to be $0.5\ dB/(MHz \cdot cm)$. Non-linearity was neglected in the described simulation.

The authors simulated this medium with the parameters of a Cephasoncis system with a 1-dimensional probe with 128 elements and a $3.75cm$ face transmitting at 5 MHz. In this work, no B-mode reconstructions of the simulations were displayed. This was likely due to the lack of realism of the simulations, which could be due to the low number of scatterers per wavelength squared or numerical artifacts in the simulation.

The aforementioned simulation parameters were used to generate a data set of 6026 training samples and 800 test samples. The networks were trained for 800 epochs on an NVIDIA GTX 1080i GPU. Before being fed to the network, the raw channel data was amplitude corrected with time-gain-compensation (TGC) at a rate of $0.25\ dB/cm$ at $1540m/s$, and the transmit pulse is removed from all data.

The language used around ultrasound channel data is not always clear. Here, the author spoke of “cropping transmit signals,” largely using computer vision vocabulary. In ultrasound beamforming, a related but separate topic from sound speed estimation, this removal of transmit signals is often referred to as “setting $t_0$”; i.e., defining the point in time relative to the beginning of a giving recording event $t_0$ at which the wave is said to be propagating through the medium and not, e.g., the transducer perfect matching layer.

On the validation set, the trained network could reconstruct a sound speed distribution. Specifically, it was stated that the reconstruction worked "well on large objects, but could miss fine details" [40]. Absolute error figures were displayed with a threshold of $50\ m/s$, which displayed overall agreement between the estimated sound speed distribution and the respective sound speed distribution used for simulating the training set. The authors reported a mean absolute error of $11.5 \pm 14.9$ on the training set and $12.5 \pm 16.1$ for the "middle network" on the test set.

Real data measurements with the simulated Cephasonics ultrasound device were also performed on polyurethane phantoms with inclusions and in-vivo samples. All real data acquisition qualitative evaluation was performed, and the results were discussed. In the polyurethane phantom, the resulting sound speed estimations could reconstruct shapes correctly, but the estimated sound speed values were off by up to $150\ m/s$. On the reconstructed distributions from in-vivo data, the resulting estimations included values outside the normal envelope of healthy tissue.

Together, this initial work showed the potential of using deep learning to estimate sound speed. The results showed promising initial findings, and the approach was very novel. Nevertheless, as with many young research fields, there was room for future research work in the space.

A new and more recent investigation on the use of deep learning for the task of ultrasound sound speed reconstruction was presented by Jush et al. [78]. In this work, they extended the network architecture of [40] to map IQ data to sound speed distribution maps. Though IQ data have advantages, in this work, the switch to IQ data was motivated by its availability on research ultrasound devices.

The k-Wave simulation suite was again used to simulate random ellipses in media. A constant background sound speed of $1535\ m/s$ and five randomly placed ellipses of a constant sound speed between $1300$ and $1700\ m/s$. This range is just slightly outside the normal envelope of $1400$ and $1600\ m/s$ for in-vivo tissue and was motivated by network generalization. A constant background density of $1020\ kg/m^3$ was set and, as in [40], a scatter distribution of two scatterers per lambda squared with $\pm 3\%$ variation in density was applied in the density channel of the medium. In total, 670 samples were generated, 6000 for the training set and 700 for the test set.

The word **scatterer** refers to a sub-wavelength particle in a medium, e.g., cells, collagen, or capillaries, that do not reflect but scatter the wave back towards the sender. The resulting received signal, when reconstructed, results in an imaging artifact called **speckle**. It does sometimes happen that these **scatterers** are referred to as **speckles**, which is not the case.

The network in this work built upon that proposed by [40] and was also based on a VGG style architecture. Aside from the use of complex data, the proposed method had one other major differentiating feature. The proposed network also had multiple encoder branches, but rather than encoding different transmissions, the network employed two encoder branches to encode

the I and Q components of the IQ data separately. In order to account for this network change, only one plane-wave transmission at $0^\circ$ steering angle was fired using the center $64$ elements of a $128$ element transducer.

Again following [40], different variations on the proposed network were compared as part of the evaluation, but only on simulated data. Of the networks they evaluated, the best was the "Cartesian" *IQ-net* with a MAE of $4.66 \pm 0.26$ on the validation set, compared to the base-line *RF-net* from [40] MAE of $5.96 \pm 0.33$.

#### Challenges for Deep Learning

Two challenges for deep learning sound speed estimation models are data collection and labeling. For supervised deep learning, there is not yet an accurate way to manually label ultrasound signals with a local sound speed label. For this reason, full-wave simulations are used to create a paired dataset of sound speed distributions and channel data. This approach is nevertheless challenged by the requirement for simulations to be parameterized accurately in order to model realistic transducer characteristics and tissue property distributions. Until now, deep learning approaches have not quantitatively proven their efficacy, and the simulations used to train such models have not accurately represented anatomical targets.

### 5.2.3 Potential of Deep Learning

Despite the aforementioned challenges in the application of deep learning approaches in ultrasound imaging with raw channel data and its derivative forms, there are still many exciting ways in which deep learning can augment current technologies in ultrasound imaging. First, the use of more realistic simulation modeling and methods can improve the data generation pipeline. These methods need to realistically model the targeted imaging medium, often human tissue and be highly accelerated and distributable parallel compute infrastructure. In order to do so, more accurate information on the properties and distribution of in-vivo tissue needs to be collected. This collection process could be achieved in highly connected point of care ultrasound devices with quantitative triggers that return channel data when a given trigger criterion is met. These criteria could be quantitative imaging measures like, e.g., speckle size, signal coherence, or a second-order measure like segmentation confidence from a neural network trained for a high-level task. Of course, a better understanding of tissue properties and their distributions could also be collected via the classical research approach for both in-vivo and excised tissue in a laboratory setting under controlled conditions. This approach has the disadvantage of scale but the potential advantage of higher and more standardized data quality. Regardless of how more advanced data and simulation quality are achieved, such advances would allow for realistic tissue simulation at a scale that mimics the distribution of real-world ultrasound applications.

**Contributions:**

- A novel approach to generating randomized and realistic simulation data in the k-Wave simulation suite
- A deep neural network (DNN) trained on beamformed IQ data generated from our proposed firing pattern to generate sound speed distributions
- First quantitative results of a DNN on phantom data consistent with traditional sound speed measurements, along with evaluation on in-vivo data as well
- Evaluation of temporal consistency that displays invariance to artifacts such as thermal noise between frames

## 5.3 Methodology

### 5.3.1 Data Pre-processing Pipeline

The raw channel data signals are converted into a representation suitable for interpretation by a DNN, and elements of realism must be included in this data in order to generalize to real-world ultrasound signals. First, the channel signals are resampled in the temporal dimension to a sampling frequency consistent with ultrasound systems. For the sake of simulation stability, as discussed in Chapter 3, the sampling frequency of a k-Wave simulation is dictated by the stability criterion of the underlying numerical formulation. Often this leads to a higher sampling frequency of simulated raw channel data. Resampling the data at train time to the target sampling frequency of the real-world transducer negates the need for up-sampling at inference time on a real machine, which could induce interpolation artifacts and decrease inference frame rate.

The simulated signals are then convolved with the transducer's impulse response. As presented in Chapter 1, the impulse response acts as a bandpass filter of broad-band wave impinging on the transducer face. Without modeling the impulse response of a given simulation, the resulting data would contain high-frequency artifacts which are both above the frequency range of the ultrasound transducer and the maximum stable frequency of the k-Wave simulation. Their presence in the raw radio frequency channel data can be attributed to simulation artifacts and safely removed to improve data realism before training a neural network.

Thermal noise augmentation (TNA) is performed by adding white thermal noise to channel data with an augmentation likelihood $p_{TNA}$. Thermal noise is an artifact resulting from electronic noise in ultrasound devices. In k-Wave simulations, this artifact is missing since it is added to the signal after the transducer and is therefore out of the scope of the simulation modeling [69]. We, therefore, add this noise back into the data to faithfully model the signal recorded by a real-world ultrasound device. Here, TNA is defined by an upper and lower bound in noise amplitude relative to the transmit signal's Root Mean Square (RMS).

This uniform distribution is randomly sampled, and the TNA is generated via the method proposed in [69]. Due to the attenuating tissue model, the constant TNA reduces SNR over depth, as is the case in real-world ultrasound imaging.

Next, a start delay is applied to the RF channel signal to correctly align a defined $t_0$ for every transmitted plane wave. As discussed in Chapter 1, $t_0$ defines a standardized point in space along the wave where $t = 0$. This aligns multiple transmissions through a medium during beamforming independently of steering and focus. The definition $t_0$ can vary between devices and simulation platforms. In this work, $t_0$ is defined as the end of the pulse of the last firing element for a given transmission event. This definition removes the transmitted pulse from the training data, as was achieved in other works through "cropping," which would otherwise introduce a large amplitude discrepancy in the training data. A large-amplitude discrepancy in training data would hinder the re-use of convolutional filters in the network on all the data and therefore lower the learning capability of the neural network and thereby impair network training [71, 93].

Next, RF signals are Hilbert transformed to generate a complex IQ representation of the data. The data of each plane wave is then beamformed individually via dynamic receive beamforming with an assumed sound speed $c_0$. This process creates complex beamformed IQ images similar to [158] which assigns a complex phase and amplitude value to a spatial location. There is, of course, a reconstruction error in the data due to the assume $c_0$, which represents the problem that is being addressed by this work. Nevertheless, the collection of signals spatially reduces the number of tasks the neural network has to perform and allows the network to compare spatial features between the three plane waves in the same spatial frame of reference.

Lastly, the complex IQ components from the same spatial location are mapped to separate channels. The final data block that is passed to the network has the dimensions [plane wave, IQ, elements, samples].

### 5.3.2 Network Architecture

We design a deep, fully convolutional neural network $F$ to take, as input, three beamformed IQ images of a medium (one for each angled plane wave transmission) and output an estimated sound speed map of the medium defined as

$$F : \mathbb{C}^{N \times M} \mapsto \mathbb{R}^{N \times M},$$

for an image size of $N \times M$ pixels. This network consists of three input dense blocks (one for each angled plane wave), a bottleneck and four decoder dense blocks that output the model sound speed estimation. The overall architecture can be seen in Figure 5.1.

Furthermore, the encoder block input accepts three complex beamformed plane waves discussed previously. The separate processing of each plane wave ensures the extraction of robust features, such as phase coherence and feature offset individually; later blocks can use that to generate an accurate sound speed map. These extract features can then be compared and further interpreted in blocks two and three to extract higher-order relationships between

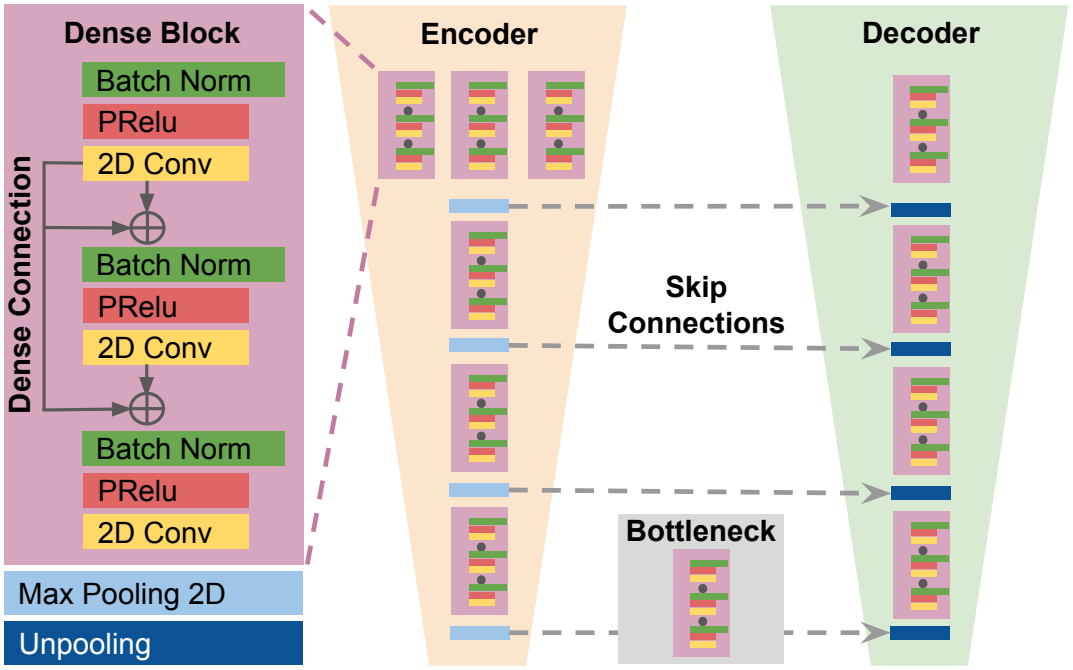



**Fig. 5.1.** Overview of the proposed architecture. Our model is composed of an encoder that individually processes three beamformed IQ images, whose features are concatenated after their individual dense blocks, a bottleneck, and a decoder that utilizes unpooling and produces the sound speed estimations. Dense skip connections are used within each dense block, and long-term skip connections are placed between encoder and decoder to enhance gradient flow and maintain feature quality.

separate image regions. These relationships can then be used to infer properties such as background sound speed. After each plane wave is passed through an individual dense block, the three plane wave features are concatenated along the channel dimension and collapsed via a $1\times1$ convolutional layer.

Our network utilizes dense network blocks [75] that incorporate dense skip connections to enhance the gradient flow and maintain feature quality. Skip connections [144] are added between each dense block of the encoder and decoder to prevent vanishing gradients and enhance network trainability. These skip connections also allow low-level, fine-grained features extracted from the first three input encoder blocks to pass them directly to the decoder block. This can enable high-frequency phase and amplitude filters to infer a high-resolution sound speed map during decoding. Together, this multi-scale network enables accurate sound speed estimates.

In this work, we replace ReLu activations with PReLu [62], which has been shown to improve model fitting and reduce the risk of overfitting. Furthermore, batch normalization layers are replaced with instance normalization [175] which has also been shown to enhance training dynamics in noise-sensitive applications. Our encoder and decoder are comprised of stacked dense blocks connected via 2D max pooling and unpooling blocks, respectively. Every dense block consists of three convolutional layers, the first two of which have a kernel size of $5\times5$ with stride one and the third one a kernel size of $1\times1$ and stride 1.

Our model is trained to estimate a target sound speed map, given three beamformed IQ images from angled plane waves as input. The Mean Square Error (MSE) is used as the loss function between the estimated and target sound speed map.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{5.13}$$

In Equation 5.13, $n$ represents the number of datapoints of a sample, $Y_i$ represents the observed values, and $\hat{Y}_i$ represents the predicted values. We use weight decay with L2 regularization is also utilized to avoid overfitting and maintain weight sparsity [49].

$$\Omega(\Theta) = \frac{1}{2}||\omega||_2^2, \tag{5.14}$$

where the regularizing term $\Omega$ is parameterized by the optimization parameters $\Theta$ and the weights $\omega \subset \Theta$. The total loss consists of a weighted summation of Equations 5.13 and 5.14.

### 5.3.3 Proposed Transmission

To find a balance between spatial medium sampling and computation cost of generating a dataset, the selected transmission protocol has to fulfill the following criteria:

- The number of transmissions should be low since the computational complexity of the simulations scales linearly with transmissions.
- The angular shift between transmissions should be small enough for spatial correlation
- The angular shift between transmissions should be large enough to gain a large sample of the spatial frequency domain known as k-space.

## 5.4 Experimental Setup

### 5.4.1 In-Silico Simulations

The in-silico tissue models described in Chapter 3 are parameterized with the values in Table 4.1. The k-Wave simulation parameters are summarized in Table 5.1. The Gaussian filter for the background generation are sized to be $x_f = y_f = 400$ pixels and the standard deviation of the filter is set to be $\sigma = 600$ pixels. The density ratio $\alpha_\rho$ is $1.5 \pm 10\%$, in other words, uniformly sampled from the set of $[1.35, 1.65]$. In doing so, we do not use constant echogenicity in the simulations. The ever-changing density ratio creates a dataset where sound speed is more statistically independent from the resulting echogenicity maps of reconstructed ultrasound B-mode images.

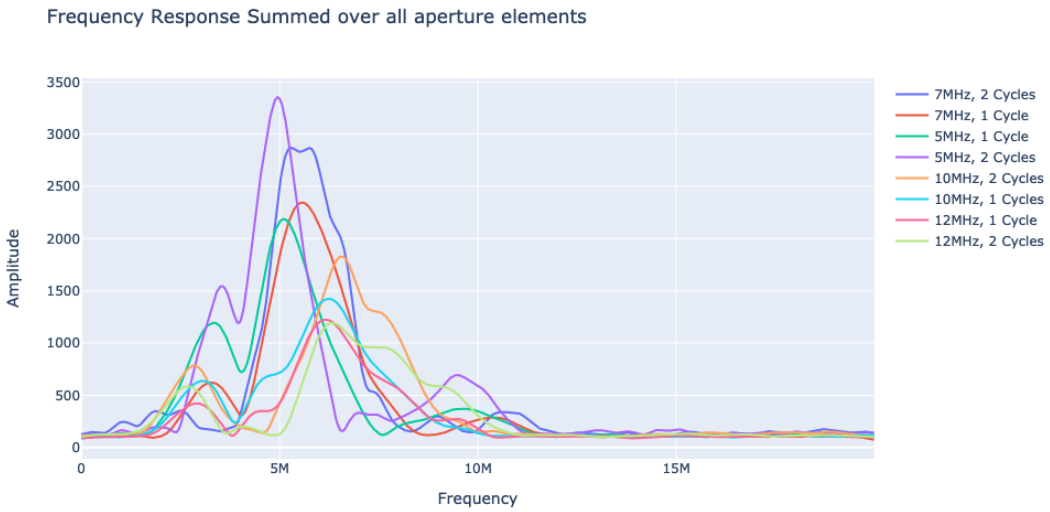



**Fig. 5.2.** Response amplitude over frequency comparison for eight different transmit configurations of the Cephasonics CPLA12875 transducer. To find the maximum sensitivity, one is interested in finding the transmit configuration with the highest response amplitude. Ideally, this response amplitude should also align with the transmit frequency, indicating a clean pulse-echo signal. In the plot above, one can see that both one cycle and two-cycle pulses were investigated. In this plot, one can see that the 5 MHz transmission with two cycles had the largest amplitude. Furthermore, all other transmit frequencies were "pulled down" by the frequency response of the transducer. This means that though a given transmit burst was transmitted at, e.g., 12 MHz, the signal received by the transducer had a peak at 6 MHz and not 12 MHz as would be expected. This indicates that the transducer does have the ability to receive at 12 MHz i.e., 12 MHz is outside of the sensitivity envelope of the transducer.

### 5.4.2 System Appraisal

The transducer modeled for the simulations is the Cephasonics CPLA12875 (Cephasonics Ultrasound Solutions, Santa Clara, California, USA) with a transmit frequency of 5 MHz, a sampling frequency of 40 MHz, and a transmit duration of one tone-burst cycle. Crucially, the transmit frequency sensitivity of the simulation was determined by empirical testing of the real-world transducer. The resulting impulse response was different than the specified 7 MHz peak sensitivity. This testing was performed by transmitting single plane waves of various configurations at a point target in water and analyzing the spectral response dependent on the transmit configuration. The results of this experimentation can be seen in Figure 5.2. The interpretation of Figure 5.2 is described in its respective caption.

Geometrically, the transducer is modeled to have 128 elements with a total aperture width of 37.5 mm, an element height of 7 mm, an element width of 0.293 mm, and a kerf, or interelement spacing, of 0 mm between elements. The transducer width and height were taken from the technical specifications of the transducer from the manufacturer. Unfortunately, the element width and kerf were beyond the scope of the specification. For the simplicity of modeling, the kerf was set to zero, which is a reasonable assumption granted that the kerf is assumed to be an order of magnitude smaller than the element width and therefore has a negligible impact on the aperture of the transducer. This modeling assumption allowed the

**Tab. 5.1.** Simulation parameters of the k-Wave simulation. All transducer properties were prepared to match the real world transducer and reduce the domain shift between the simulation and the real world.

| Property | Value |
|---|---|
| Transmit Frequency | 5 Mhz ± 10% |
| Center Frequency | 5 Mhz |
| Density Ratio | 1.5% ± 10% |
| Alpha Power | 0.75 dB/MHz cm |
| Apha Coeff | 1.5 |
| B/A | 6 |
| Bandwidth | 60% |
| Tone Burst Cycles | 1 |
| Sampling Frequency | 87.6 Mhz |
| # Elements | 128 elements |
| Pitch | 292 $\mu$m |
| Kerf | 0 $\mu$m |

simulation of the transducer with a coarser mesh grid, given that the gird resolution was not required to be small enough to resolve the element kerf.

A medium sound speed of 1540 m/s is used in order to calculate the transmit delays for steering angles of -8, 0, and 8 degrees for each plane wave transmission, respectively.

Steering delays are today defined by their propagation angle [112]. Unfortunately, the real propagation angle of the wave in the interrogated medium differs based on the sound speed of the medium. When the transmitted wave impinges on the medium, the sound speed modifies the steering angle in accordance with snell's law and requires correction terms in the reconstruction of coherently compounded B-mode images. A more apt way of defining steering angles would be with relative temporal transmit offsets between elements, e.g., $x \ \frac{\mu s}{element}$, but this is not yet standard convention.

Critically, in contrast to [40], all three transmissions are simulated from the same aperture of the center 64 elements, as is more commonly used in plane wave imaging pulse sequences that utilize coherent compounding. This convention keeps the aperture of the transmission and receive the same and more easily allows for quantitative comparison of the returning signals. On both transmit and receive, rectangular apodization is employed. By using rectangular or constant apodization, the amplitude of element signals is kept constant in order to reduce implicitly added biases with more complex apodization methods. Since the task at hand is not specifically a reconstruction task, and the simulated signals are subsequently Hilbert transformed, it is necessary to consider apodization for both transmit and receive while laying out the simulation pipeline.

The medium dimensions in grid points are $N_x = 548$, $N_y = 648$ and $N_z = 126$ with a grid spacing of 58.594 $\mu$m in all directions. Constant grid spacing simplifies subsequent calculations and improves the numerical stability of the simulation [171]. The total dimensions $(x_d, y_d, z_d)$ of the simulated domain are 32 mm × 38 mm × 7.4 mm. A Perfectly Matched Layer (PML) of

size $7 \times 17 \times 9$ grid points are added to the medium to prevent signal wraparound [171]. The modeled transducer is centered upon the phantom grid.

In total, 5996 samples consisting of three plane wave simulations are generated using the k-Wave Toolbox [171], and the C++ accelerated binary on an NVIDIA Quadro RTX 6000 GPU with 64 CPU threads. The sampling frequency of the simulation dictated by the k-Wave simulation time-step is 87.6 MHz with a resulting maximum supported frequency of 12.26 MHz [171]. The GPU run-time per simulation is 620 seconds, and 43 days 38 minutes and 40 seconds for the entire dataset.

### 5.4.3 Data Processing Parameters

The simulated channel data is resampled from 87.6 MHz to 40 MHz. A Gaussian band-pass filter centered at 5 MHz with 60 % fractional bandwidth is applied to model the transducer's impulse response. TNA was performed with an magnitude range from -120 dB to -80 dB and an augmentation likelihood of $p_{TNA} = 20\%$. A $t_0$ was set to $2.75$ $\mu$s for the center transmission and $5.0\mu$s for the $\pm 8$ degree plane waves. Afterward, the data was Hilbert transformed to generate the analytical signal, and the complex components were decomposed into separate channels. The 20% likelihood was imperially chosen to allow the network to see "clean" data and be challenged by added noise in the signals.

### 5.4.4 Network Training

Our deep model is trained with a batch size of 6 for 138 epochs and a learning rate of 0.001. Validation loss-based early stopping is employed to terminate training. The Adam optimizer [85] is used with weight decay, Equation 5.14 activated with a decay rate of $e^{-4}$. The network is created in Python using the PyTorch Library v1.7 [127] and the Pytorch Lightning framework v1.2.10. Weights and Biases are used for tracking experimental metrics and figures. Our models are trained on an NVIDIA Quadro RTX 6000 GPU.

### 5.4.5 Simulation Evaluation

Our model is evaluated on a simulated validation set of 514 samples equally drawn from all classes that are meant to accurately represent the training data set yet be unseen by the network during the training stage. We report the mean absolute error between the predicted and target sound speed for each class. Furthermore, to showcase the advantage of TNA, we compare the error distributions over all classes for two otherwise identical models, trained with and without TNA. Lastly, we further investigate the effect of thermal noise on the models by comparing the error over depth for three levels of additive thermal noise and our baseline without noise.

### 5.4.6 Phantom and In-Vivo Evaluation

To evaluate the predictive efficacy of our model, phantom and in-vivo studies are performed using a Cephasonics Griffin with 64 channels and a CPLA12875 transducer (Cephasonics Ultrasound Solutions, Santa Clara, California, USA).

The sound speed of a homogeneous CIRS Phantom Model 040GSE (CIRS Inc, Norfolk, VA USA) is verified via speckle brightness [120] to be 1558 m/s. The technical specification of the phantom specifies the phantom to have a sound speed of 1540 m/s, which was not reproduced in the speckle brightness measurement. The age and condition of the phantom are assumed to be responsible for the deviation in sound speed from its factory specification.

Next, a bovine steak is prepared, and its sound speed is measured to be 1566 m/s in a distilled water bath (24.6°C, 1495.8 m/s [107]) using the method described in [88]. A water bath of distilled water at room temperature (24.6°C) is placed in a ceramic vessel. The sound speed of the distilled water is measured to be 1495.8 m/s using the method described in [107]. The transducer is mounted on a stand and placed approximately 2.5 cm from the base of the vessel. Two ultrasound measurements were performed, whereby once an empty measurement of water was taken and once the steak was inserted into the water bath between the transducer and the reflective bottom of the vessel. The sound speed of the steak is found via the insertion method [88] to be 1566 m/s.

To test the real-world performance of the trained estimator, the steak is cut in two separate slices of 8 mm and 4 mm and stacked on the CIRS phantom as a two-layer model. The regional mean sound speed error is estimated for regions of interest (ROI) in the steak and at proximal and distal locations in the CIRS phantom. The differentiation of proximal and distal regions is performed to showcase the effect of depth-dependent SNR on the model predictions. The setup was undertaken to evaluate the hypothesized mechanism of degradation of prediction quality over depth due to lower SNR in the signal, i.e., the proportionally greater amount of thermal noise over depth.

Furthermore, to reduce selection bias and evaluate the temporal consistency of our model, the regional sound speed estimates are averaged over 100 consecutive static frame measurements. This experimental setup, adopted for the first time in ultrasound sound speed estimation, allowed us to quantify the influence of the thermal noise and other error factors on the sound speed estimations of independent measurements. Sound speed estimation methods have historically been evaluated based on their bias and precision and on homogeneous sound speed phantoms. Though insightful, clinical B-modes more often scan regions of strongly heterogeneous tissues with varying scatterer densities, attenuation rates, and sound speeds. Furthermore, thermal noise from the ultrasound scanner can further corrupt the signal from a given interrogation. Therefore, it is important to evaluate sound speed estimation methods in realistic settings of homogeneous tissue distributions with regional error values for known sound speeds. Since multiple samples are usually collected in the process of evaluating a method, this can lead to either conscious or unconscious bias in the evaluation of the method. In order to reduce this, a random selection of 100 continuous frames was made with which the evaluation was conducted.


**Fig. 5.3.** The experimental setup of the CIRS phantom acquisition can be seen above. A porcine steak was placed between the transducer face and the CIRS calibration phantom to serve as an aberration screen.

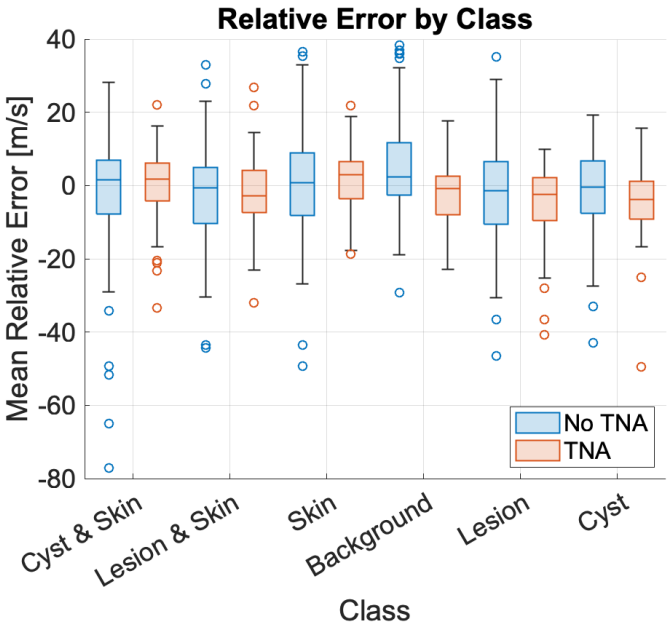



**Fig. 5.4.** Boxplot comparing sound speed relative estimation error distributions per class for the simulated validation set. The central mark on the box indicates the median value, and the top and bottom edges of the box indicate interquartile range. The black whiskers indicate the extent of the distribution without the outliers, which are denoted by circles on the plot. Overall, the model trained with TNA achieves lower relative error standard deviation and fewer outliers for all classes.

In-vivo imaging is performed on the left breast of a healthy volunteer (Age: 28, BMI: 22.4) in three regions. The volunteer was selected under an approved IRB protocol of the Technical University of Munich and provided written informed consent. Channel data for each region are acquired with the same configuration as the phantom experiments.

## 5.5 Results

The following section presents the experimental results of the sound speed estimation experiments on simulation, phantom, and in-vivo channel data.

### 5.5.1 Validation Set Evaluation

Table 5.2 displays the class-wise validation set MAE for a base network trained without TNA and one trained with TNA. The class-wise MAE is low for both models, ranging from 8.50 m/s for the TNA skin class to 16.4 m/s for the base lesion class. These MAE are small relative to the wide sound speed ranges the model is trained on. Table 5.2 also highlights that TNA substantially improves estimation error across the classes by 2.2 m/s for the cyst class to 5.5 m/s for the skin and cyst class. Overall, both the error and standard deviation are also reduced on the validation set with augmentation of TNA.

Figure 5.4 shows the relative average error for each class given models trained with and without TNA. Though the performance of both DNNs is acceptable, it is clear TNA contributes

Tab. 5.2. Sound speed estimation MAE and standard deviation per class for models trained with and without TNA. Estimations are shown in m/s.

| Class | No TNA | TNA |
|---|---|---|
| Cyst & Skin | $16.1 \pm 12.4$ | $10.6 \pm 5.10$ |
| Lesion & Skin | $15.5 \pm 8.70$ | $12.0 \pm 5.70$ |
| Skin | $12.9 \pm 9.10$ | $8.50 \pm 4.00$ |
| Background | $12.8 \pm 8.84$ | $7.90 \pm 3.70$ |
| Lesion | $16.4 \pm 8.70$ | $12.7 \pm 7.30$ |
| Cyst | $12.8 \pm 6.30$ | $10.6 \pm 5.90$ |
| Overall | $14.3 \pm 9.20$ | $10.3 \pm 5.60$ |

towards reducing the standard deviation (signified by box size) of the relative error and number of outliers (signified by circles).

### Effect of Thermal Noise over Depth

In Figure 5.5, we show the effect of additive thermal noise over depth on the predictions of networks trained with and without TNA. This comparison should show how the addition of TNA augmentation benefits training neural networks for ultrasound applications. By evaluating the sound speed estimation over depth, we hope to show the effect that a low SNR due to error terms such as thermal noise in the raw channel data has on the estimation quality. This understanding would better help those looking to control for the potential shortcomings of the proposed method.

We evaluate three scales of additive noise, specifically $-80$ dB, $-100$ dB, and $-120$ dB relative to the transmit signal RMS, along with a baseline measurement without noise. First, it can be seen that the network trained with TNA (Bottom) is robust to thermal noise since the error remains low over the entire depth of the measurements. The network trained without TNA (Top) is severely affected by thermal noise present in the channel signals for all noise levels, with increasing error from -120 dB to -80 dB. As the signal weakens due to attenuation, the constant thermal noise reduces the SNR and weakens the signal. For the network trained without TNA, this low SNR reduces the estimation quality. The baseline sound speed error shows that the network trained without TNA can accurately estimate the sound speed over depth on the validation set when no noise is added. For noise levels -120 and -100 dB, the model trained without TNA underestimates the sound speed in the medium. For the noise level -80 dB, the model underestimates to a depth of 1.6 mm and then overestimates the sound speed in the medium. Moreover, the effects of thermal noise get more prominent deeper in the image; for the network trained with TNA, the estimation marginally worsens after 2 mm, while for the one. The network trained with TNA only marginally underestimates the medium sound speed after 2 mm depth. Hence, there is a clear relationship between the SNR of the signal and the model prediction, and the utilized TNA is beneficial.

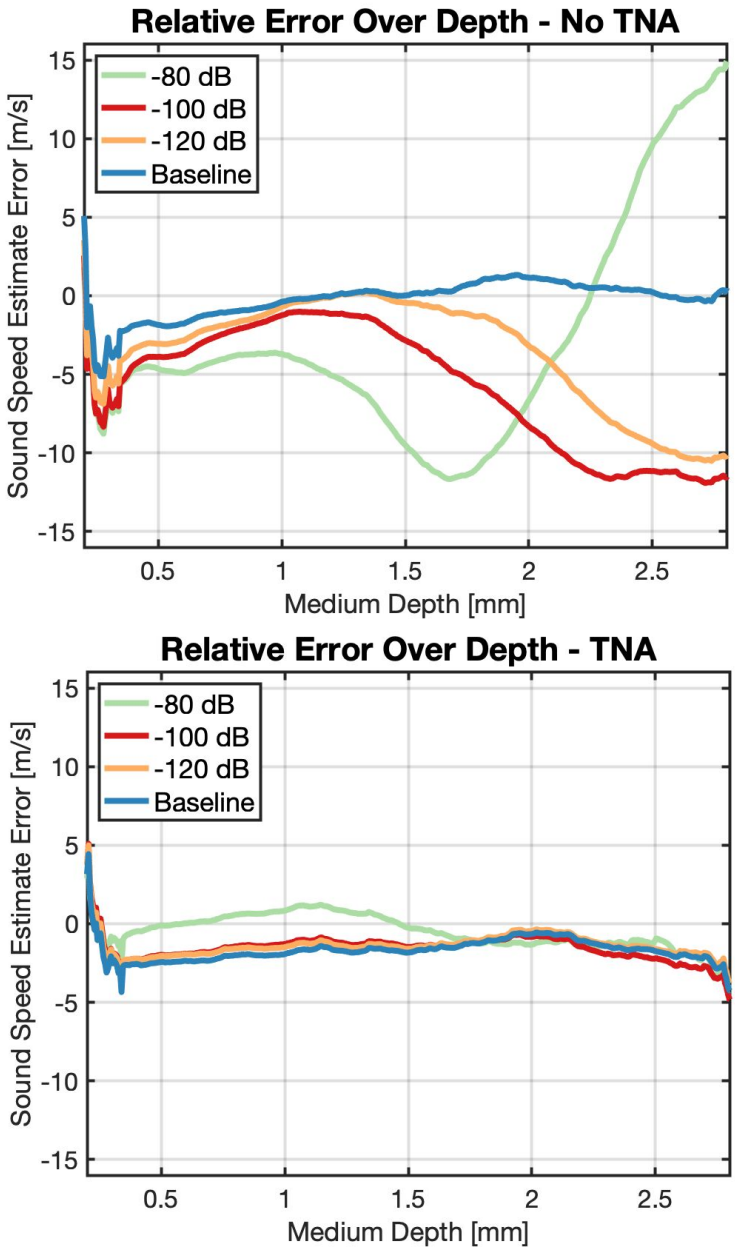



**Fig. 5.5.** Relative sound speed estimation error over depth for the simulated validation set for three levels for the addition of thermal noise and baselines without added noise. Our model trained with TNA (Bottom) is markedly more robust to the addition of thermal noise, while the one trained without TNA (Top) appears to be sensitive to noise. Due to this fact, its performance decreases proportionally to the decrease of SNR over the depth of the image.

### Qualitative Evaluation

Qualitative results of simulated B-modes for all classes and their respective sound speed estimates are shown in Figure 5.6. Our proposed simulation pipeline creates B-mode images with an overall realistic breast-tissue appearance. A realistic appearance of a simulated B-mode image is a strong initial indicator of simulation quality.

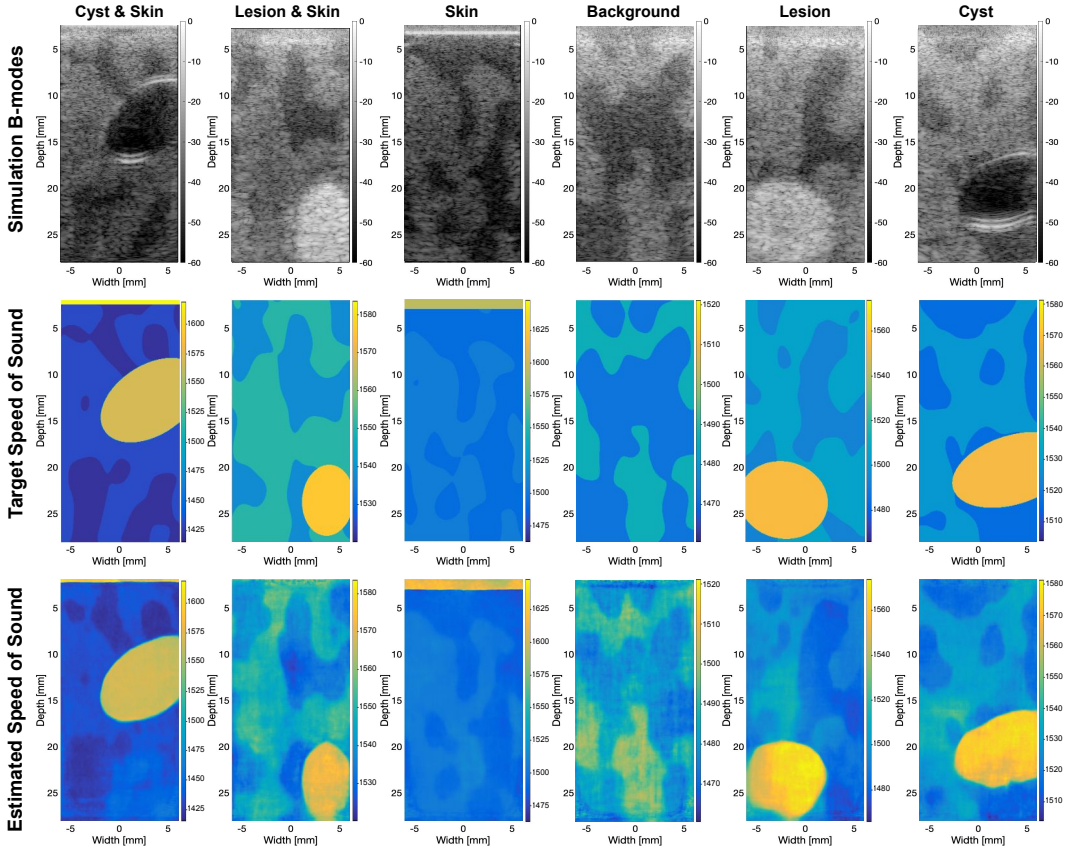



**Fig. 5.6.** Simulated B-modes from six classes with target sound speed and model estimation. Our simulations produce realistic B-mode images, and our model successfully estimates sound speeds and contours of cysts, lesions, skin, and background.

Consistent with the results shown above, our model is able to successfully estimate sound speed distributions throughout the simulated domain for all data classes. Contours of both anechoic (dark) and echogenic (bright) features in the images are recovered. Also, the sound speeds within simulated anatomical regions are estimated with a low error. In B-modes of classes with cysts, reverberation artifacts are often present on the boundaries of the simulated cysts, where a high sound speed boundary gradient is present. Reverberation artifacts are furthermore common in cysts since the propagated wave can become "trapped" in the cyst, reflecting back and forth, and therefore has a much longer travel time when returning to the transducer (see Chapter 1 for more information on reverberation). Nonetheless, the model is able to successfully generate accurate sound speed estimates within cysts. The skin and background regions are also consistently delineated from their surroundings.

### 5.5.2 CIRS Phantom Evaluation

To evaluate the generalization ability of our model, we evaluate its performance on a layered phantom that was not represented in the training set. The decision to evaluate out of distribution samples with a real-world transducer was made in order to stress test the performance boundaries of the model. One hundred ultrasound frames are acquired with a real transducer, and the results for the layered phantoms with both a 4 mm and 8 mm bovine steak phantom

are shown in Table 5.3. As stated in Section 5.4.6, the measured sound speed of the steak is 1566 m/s, and that of the CIRS phantom is 1558 m/s. The evaluation of the model estimation is performed in three discrete ROIs that extend across the full image aperture and are depicted by red, yellow and green boxes in Figure 5.7 in order to evaluate the influence of thermal noise and attenuation along with other real-world factors on the model's estimation at varying depths. Again, the evaluation of sound speed estimation over depth aims to show the effect that a low SNR has in estimation with raw channel data. This experiment also serves as a comparison to the performance over depth as evaluated in Section 5.5.1. A correlation in the results of both of these sections would serve to further strengthen the confidence that the simulations are realistic enough to accurately train a model.

**Tab. 5.3.** Sound speed estimations and errors for the CIRS and steak phantom predictions compared with the insertion and speckle brightness methods in m/s. Estimations, errors, and standard deviations are computed over 100 consecutive frames.

| | Traditional Measurements (m/s) | 4mm Steak | | 8mm Steak | |
|---|---|---|---|---|---|
| | | Estimation | Error | Estimation | Error |
| **Steak (red)** | 1566 | $1564.4 \pm 3.60$ | $-1.60 \pm 3.60$ | $1564.6 \pm 2.70$ | $-1.40 \pm 2.70$ |
| **CIRS Background (yellow)** | 1558 | $1555.6 \pm 4.43$ | $-2.40 \pm 4.43$ | $1558.9 \pm 2.49$ | $+0.90 \pm 2.50$ |
| **CIRS Background (green)** | 1558 | $1544.7 \pm 6.90$ | $-13.9 \pm 6.90$ | $1542.7 \pm 4.80$ | $-15.3 \pm 4.80$ |

Even though our model is solely trained on simulated ultrasound signals, it is still able to successfully infer the sound speed of these two-layered phantoms in agreement to ex-vivo sound speed measurement. The mean error for the steak layers ranges from 1.6 m/s for the 4 mm steak to 1.4 m/s for the 8 mm one. Furthermore, the sound speed for the top ROI of the CIRS phantom is also successfully estimated with a mean error of 2.4 m/s for the 4 mm steak and 0.9 m/s for the 8 mm one. As can be seen in Table 5.3 the standard deviation values of the estimations among the 100 consecutive frames are also low, ranging from 2.7 m/s to 6.9 m/s, showcasing the temporal consistency of our model predictions for real-world data.

Finally, we can see that the predictions for the bottom 2.9 mm of the CIRS phantom (green ROI) have larger error than those of the top, ranging from 13.9 m/s for the 4 mm steak phantom to 15.3 m/s for the 8 mm one. The total range of the 8 mm phantom is 1516.2-1653.7 m/s and 1518.9-1645.9 m/s for the 4 mm phantom. The prediction of the model trained without TNA for the bottom region of the CIRS phantom is 1536.7 m/s $\pm$ 4.72 m/s for the 4 mm steak phantom and 1510.7 m/s $\pm$ 8.11 m/s for the 8 mm one. This shows the superiority of the model trained with TNA, which decreases the error from 29.3 m/s to 13.9 m/s for the 4 mm steak phantom and from 55.3 m/s to 15.3 m/s for the 8 mm one. These results are in line with those on the validation set and are promising for the generalization of our trained model beyond simulations to out-of-distribution heterogeneous tissues.

### 5.5.3 In-vivo Evaluation

Figure 5.8 shows the predictions of our model for three breast regions in a healthy volunteer. As with the phantom evaluation, we calculate the average sound speed over 100 consecutive frames with a static probe for the in-vivo measurements. With no specific ROI or ground

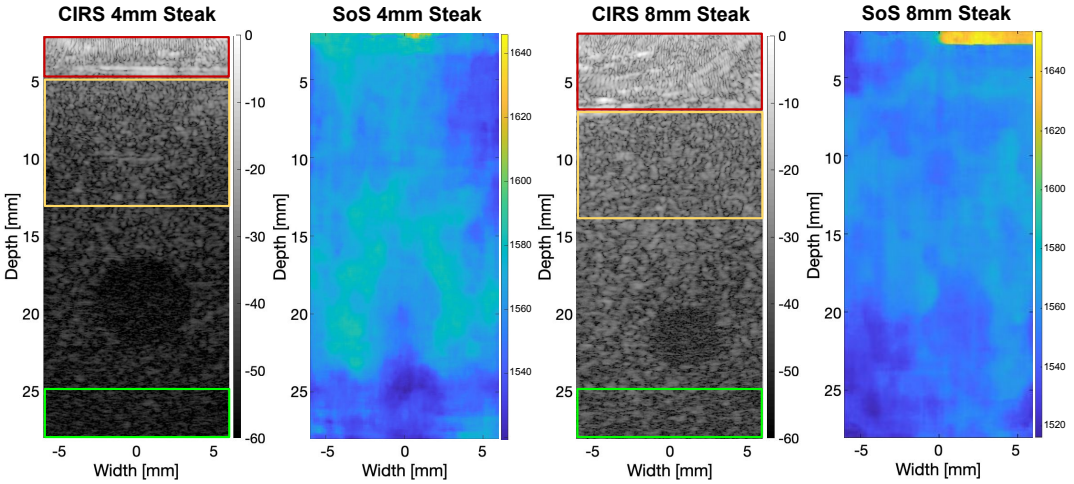



**Fig. 5.7.** Sound speed estimations for CIRS and steak layered phantoms along with B-mode images. The red ROI delineates the steak at a depth of 4 mm and 8 mm respectively. The yellow ROI delineates the top of the abutting CIRS layer and is 8.6 mm thick (left) and 6.6 mm thick (right). A green ROI encloses the bottom 2.9 mm of both phantoms. Model estimations are coherent and agree with the measured sound speed of 1566 m/s for the steak and 1558 m/s for the CIRS background.

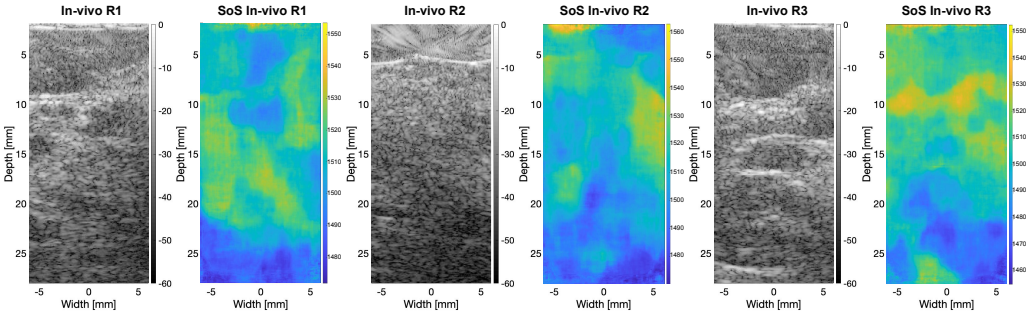



**Fig. 5.8.** In-vivo sound speed estimations along with B-mode images for three breast regions R1-R3. Our model can estimate coherent maps for all three breast regions, with breast gland sound speed values within the sound speed range measured in [5, 59, 118]. Moreover, tissue contours around fat and connective tissue are also correctly delineated by our model.

truth, the estimated sound speed of the entire field view is evaluated. The overall mean sound speed over 100 frames for R1, R2, and R3 are 1518.0±5.3, 1500.1±6.1, and 1499.0±3.4 m/s, respectively. These values are consistent with each other, and the literature on the sound speed of measured glandular breast tissue of 1505.0±47.3 m/s from the Foundation for Research on Information Technologies in Society (IT'IS) [59], and 1510 m/s from Nebeker et al. [118]. Also, the model predictions align with the values of our simulated dataset, where breast gland is modeled with sound speed between 1480 m/s and 1528 m/s following [5].

## 5.6 Discussion

Our proposed simulation method for breast modeling creates realistic breast tissue US B-mode images, and our data processing pipeline successfully converts the simulated ultrasound data into a suitable representation for training a deep learning model for sound speed estimation.

The realism of our simulation is, as previously stated, the first indication of good domain fit, i.e., the data simulated represents the true data distribution well. This can be seen in the expected contrast distributions, the realistic ultrasound artifacts such as gross and local reverb, shadowing, and amplification. Lastly, the reconstruction shows small, tight speckle kernels, indicating a high-quality reconstruction resolution and good spatial correlation between plane wave transmits.

An advantage of the complex spatial IQ representation is that our model can process both the magnitude and phase information when predicting spatial sound speed distributions. This representation allows the model to take into account phase shifts with ease. Would one train a network on raw radio frequency channel data, one would expect the network to learn phase and amplitude features in order to solve the given task and then spatially correlate signals in different temporal regions of the signal to its respective spatial origin. By pre-processing the data, the network can use these features without having to learn filters to extract them within the network, e.g., the phase shift.

It has been shown that networks similar to our proposed fully-convolutional architecture take a multi-scale context into account [49, 114]. Both large anatomical features, as well as local phase-shift features, are taken into account when generating sound speed estimates. This is enabled by the high-frequency filters of shallow layers and larger receptive fields in deeper layers of the proposed architecture [49] which are collected via skip connections and combined with the decoder weights to generate the final sound speed estimates. These architectural decisions make the proposed architecture a good candidate for sound speed estimation, as shown by the presented results.

Due to the constant sound speed assumption used for beamforming, the geometry of B-mode images can be spatially distorted compared to the geometrical layout of their respective medium. This general property of B-mode images is also present in the reconstructions of the simulated B-modes where the B-mode reconstruction does not spatially correlate one-to-one with the simulation medium. This is especially prominent below the lesion regions for the classes Lesion & Skin and Lesion in Figure 5.6 where the lower lesion boundary is not pictured in the B-mode but is visible in the sound speed simulation medium. From the B-mode alone, one can infer that the sound speed of the lesion is much lower than the background sound speed, and therefore the wave propagation through the entire medium takes longer than the assumed sound speed would expect. The delay in the wave propagation leads to anatomies that are spatially closer to the transducer in the simulation medium, appearing to be lower in the B-mode. With knowledge of the layout of the prior, relative sound speed estimations can be made by a trained eye.

It is, therefore, interesting to examine the estimations produced by the network and try to understand which features are being used for sound speed estimation. Echogenicity, i.e., the brightness or darkness of a region, can be indications of local sound speed variations between media and the standard deviation of the sound speed of the scatterers within the medium [33, 64]. Notably, the estimation of the sound speed maps does not simply correlate with the echogenicity in the B-mode images as might be expected. The estimated sound speed maps correspond correctly to the spatial distribution of the target sound speed maps and not the B-modes. This indicates that signal echogenicity is not the only feature used for

sound speed estimation. Therefore, one can infer that both the relative spatial positioning of geometrical features and local phase shifts and echogenicity are all being taken into account by the network when estimating sound speed.

The real-world predictions of the layered CIRS and steak phantoms in Figure 5.7 show that both cases are consistent with a homogeneous background, which was unseen in the training set. The sound speed difference between the steak and CIRS layers is measured with the insertion method to be eight m/s. Our model estimates a sound speed with an accuracy of 8.8 m/s for the 4 mm steak phantom and 5.7 m/m for the 8 mm phantom. These results are close to the measurements of the insertion and speckle brightness methods. Furthermore, given the homogeneous medium, the network did not infer a sound speed distribution with the appearance of the breast tissue from the training set but rather correctly inferred a homogeneous sound speed. This indicates that the network has learned a robust collection of features that allow it to generalize beyond the training data and that these features also apply to real transducer data and out-of-distribution property geometries. It would normally be expected that the performance of a network would deteriorate given out of distribution samples. Nevertheless, the macro sound speed estimate is accurate even for out-of-distribution homogeneous samples.

Two regions of over-estimation (1620 m/s) can be seen in the top 1-2 mm of both phantoms. This kind of overestimation is especially visible in the 8 mm steak phantom and resembles the skin class from the training set. On the one hand, this might be expected as the network could be expected to extract features that correlate to the top of the image and, therefore, with the presence of skin in the B-mode. Yet, when median absolute distance outlier removal is applied frame-wise to the sound speed in the region of interest, the sound speed estimate is $1562.3 \pm 2.6$ over 100 frames, only modestly increasing the regional error by 2.3 m/s. It is possible that the high sound speed estimate in the steak region delineates a region of high sound speed tendon or connective tissue. During acquisition, neither local sound speed estimation nor a high-resolution tomographic reconstruction of the tissue was performed. It is further possible that the "high sound speed region" was added in an upper layer due to an indicative feature of aberration lower in the network. Though the network reasoning is currently not interpretable, there are multiple possible reasons for the presence of this region.

The in-vivo evaluation showed global sound speed estimates in line with reference values from the literature. Unlike the homogeneous phantom models, the in-vivo estimates displayed the expected tissue variation in the sound speed estimate, which resembles the underlying breast tissue distribution. Again this is an encouraging result, indicating that the network can differentiate subtle differences in tissue sound speed at inference time. Furthermore, all estimated values in the in-vivo estimation were within the expected range for in-vivo breast tissue.

While physics-based models are often limited to correlating sound speed with spatially local features, convolutional neural networks can consider the global spatially distributed features when estimating sound speed. Specifically, physics-based models often struggle to make accurate sound speed estimates in the first 5-10 mm of a scan due to a multitude of complicating factors, such as lack of wave formation, contact interfaces, and limited angular sensitivity, that

can invalidate the underlying assumptions upon which the model is based [73, 158]. Since neural networks model the training data and not a canonical model, it can be hypothesized that spatial aberration relationships can be used as features for sound speed prediction, e.g., the spatial coherence in the middle of the image can be interpreted by the network as an indication of the sound speed of the abberating medium above.

The slight sound speed underestimation in the bottom of both the phantom and in-vivo scans could be attributed to the lower SNR deeper in the medium. Thermal noise present in real-world transducers and TNA contributes towards bridging the performance gap but still does not completely alleviate the problem. Further, the thermal noise amplitude used in the TNA might not directly match the amplitude in the US device, in part due to mismatched attenuation values. Other sources of noise in the signals from the lower regions could lower the SNR and contribute to the performance loss. Methods of increasing SNR such as higher angular sampling frequency by an increased number of plane wave firings could potentially alleviate this problem and lead to improved performance and greater scanning depths. These improvements would, of course, come at a computational cost when generating simulation data. This was the original reason why these were not performed in the scope of this work.

In the development of this model, it became apparent that the distribution shift between simulations and real data is one of the greatest challenges in generalizing deep neural networks trained on simulations for real-world usage. A data distribution shift can be caused by, among other factors, transducer-specific accidental signal encodings such as cross-talk, greater variety in the spatial distribution of tissue and acoustic properties, and a more general anisotropic reflectivity of echogenic interfaces. It is, therefore, very important to correctly and accurately parameterize simulation parameters in order to ensure the accuracy of the resulting radio frequency data.

Our phantom and in-vivo results display the proposed method's robustness by correctly predicting sound speed on out-of-distribution data and under the influence of real-world factors. This can be attributed to the proposed anatomically realistic simulations and the data pre-processing pipeline with TNA that improve generalization to real-world signals.

Furthermore, the robust evaluation of our method goes beyond the standard protocol for deep model evaluation in medical imaging. This evaluation pipeline includes testing our model on external data sources that were not included in the training distribution and reporting the model predictions over 100 US sequential frames. The low standard deviation of our errors shows the stability of our predictions over 100 consecutive frames. This approach could set a new precedence for the evaluation of the consistency of sound speed estimation for both physics-based and deep learning models.

Future work includes more realistic modeling of real transducers and in-vivo artifacts. The dataset could be further extended to include irregularly shaped lesions to model malignant tissue with irregular boundaries. Such modeling will be crucial for the development of robust and generalizable sound speed estimation models with DNNs. Furthermore, the presented method utilizes three plane waves, which reduces the SNR of the signal at both training and inference time. It is expected that with the simulation of more plane waves to a comparable number to [158], the performance could increase further along with the computational cost.

Finally, our dataset could be used as a benchmark for sound speed estimation methods to increase their comparability, similarly to challenges in beamforming such as PICMUS and CUBDL [7, 95].

# Part IV

Conclusion

# 6 Conclusion

**Contents**



This dissertation covered the diverse fundamental preliminaries required for sound speed estimation with deep learning. These spanned the basics of ultrasound physics, the intricacies of ultrasound hardware, electronics and signal processing, and the new and exciting field of deep learning.

## 6.1 Ultrasound Fundamentals

Chapter 1 described the physical property priors required to understand the data structure of raw ultrasound data and the challenges and physical concepts that affect the data quality. From the basics of the wave equation to the concepts of attenuation, absorption, non-linearity, and reflection, basic concepts were described and explored. Furthermore, a statistical view of the concept of scattering using a Monte Carlo-based random walk approach was presented and discussed.

The physics discussed in Chapter 1 is critical for a well-founded understanding of ultrasound channel data and subsequently used data structures. The traits that differentiate ultrasound signal processing from natural imaging include the complex oscillatory nature of channel data due to the carrier frequency used for transmission, the strong signal degradation due to attenuation, scattering, and absorption and the large wavelength of the transmitted signals relative to the medium being imaged. These fundamental differences affects the way one can process the data and realistically use it subsequently with deep learning methods.

Further and continued integration of fundamental physical priors into ultrasound imaging will strengthen the modality and improve the clinical relevance and diagnostic applicability of the modality. This can only come from foundational wave physics research, especially in the field of sub-wavelength scattering modeling. The work presented here is a first step in bridging the gap between computer-aided medical procedure applications and the physical phenomena by which the images are generated.

## 6.2 Realistic and accurate simulations of breast ultrasound

Chapter 2 discussed the methods to parameterize and generate accurate and robust in-silico ultrasound simulations. Being able to represent the natural processes of ultrasound wave propagation and sub-wavelength scattering in a mathematical representation that can be stably and reliably executed on consumer-grade hardware is critical for a physically informed neural network system.

We showed that a realistic in-silico phantom can be prepared as the input for an accurate numerical ultrasound simulation with the tissue property values obtained from literature. By beamforming, the qualitative analysis and results of the simulated signals confirmed their realism and similarity with in-vivo data. The proposed simulation method in this work was able to achieve a level of simulation realism for subsequent methods to accurately infer tissue sound speed on phantom and in-vivo data. This method of simulation can be built upon for applications in other quantitative ultrasound tasks, such as attenuation or non-linearity imaging.

## 6.3 Method for the estimation of sound speed in breast ultrasound

Chapter 3 presented a new and novel method for sound speed estimation in clinical breast ultrasound images. This method built on the physical fundamentals and simulation techniques of the previous two chapters and deep learning. Furthermore, realistic data augmentation with TNA ensured that networks trained on simulated in-silico ultrasound data could be agnostic to the distribution shift to real ultrasound data. Here again, this novel approach is applicable to other applications and can be thought of as a general ultrasound signal augmentation when training deep neural networks for ultrasound tasks.

We showed the effective use of DNNs to estimate sound speed maps in both phantom and real-world volunteer data. All real-world evaluations were performed on 100 sequential frames to reduce the likelihood of selection bias and allow for calculating robust statistics of the model under noisy real-world conditions. This novel evaluation technique improved the reliability and interpretability of our results and could become a standard evaluation technique for ultrasound reconstruction and estimation algorithms. Sound speed estimation can be further evaluated in clinical settings as a potential feature for breast lesion classification.

## 6.4 Future Outlooks of Modern and Quantitative Physics-Informed Ultrasound

Ultrasound is an imaging modality that has a bright future. Based on the work presented here and other developments in the field, iterative adaptive quantitative imaging will become a reality in medical ultrasound. Deep learning has shown promise in this, and prior works, to be able to learn how to perceive quantitative property distributions that can be used in subsequent image reconstruction to improve imaging quality. Also, the use of differentiable physical constraints on deep learning models, such as Physics-informed neural networks (PINNs), can aid in accelerating the training of deep learning models [137, 138, 139]. With a differentiable physical model integrated into the training process, the data generation and training workflows can both be run at train time [97]. This union could reduce the overhead burden of generating simulation media, simulating a pre-defined wave propagation, and training the network separately. Such approaches have already been explored for the simulation of differentiable wave optics for microscope and calibration [10, 38, 123, 189]. These works aimed at classifying the intrinsics of a lens given a set of images created by a lens. Similar methods could be applied to classify the intrinsics of the medium being imaged by an ultrasound transducer by incorporating the fundamental known physical principles of wave propagation into a deep learning model. These steps outline some of the potentials that deep learning can contribute towards improving ultrasound imaging quality and diagnostic potency.

I hope that this dissertation will inspire the exploration into further research in inferring quantitative tissue properties with signal priors using deep learning. If the reader has gotten this far and still has questions on the topic or would like to discuss the work presented here, please reach out to the author. Now that this is done, let's get back to work.

# Part V

Appendix

# A

# Authored and Co-authored Publications

## Authored

1. **W. A. Simson**, M. Paschali, V. Sideri-Lampretsa, N. Navab, J.J. Dahl. *"Investigating Single-Sided Sound Speed Estimation in Breast Ultrasound with Deep Learning."*, 2021 (Under Submission)

2. **W. A. Simson**, R. Göbl, M. Paschali, M. Krönke, K. Scheidhauer, W. Weber, N. Navab. *"End-to-End Learning-Based Ultrasound Reconstruction."* arXiv preprint arXiv/1904.04696, 2019

3. **W. A. Simson**, M. Paschali, N. Navab, G. Zahnd. *"Deep learning beamforming for sub-sampled ultrasound data."* IEEE International Ultrasonics Symposium (IUS), Kobe, 2018

## Co-authored

1. V. Sutedjo, M. Tirindelli, C. Eilers, **W. Simson**, B. Busam, N. Navab. *"Acoustic Shadowing Aware Robotic Ultrasound: Lighting up the Dark."* IEEE Robotics and Automation Letters, 2022

2. M. Tirindelli*, C. Eilers*, **W. A. Simson**, M. Paschali, M.F. Azampour, N. Navab. *"Rethinking Ultrasound Augmentation: A Physics-Inspired Approach."* International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Strasbourg, 2021 (Equal Contribution)

3. Z. Jiang, M. Grimm, M. Zhou, J. Esteban, **W. A. Simson**, G. Zahnd, N. Navab. *"Automatic Normal Positioning of Robotic Ultrasound Probe based only on Confidence Map Optimization and Force Measurement."* IEEE Robotics and Automation Letters (RAL), 2020

4. T. Czempiel, M. Paschali, M. Keicher, **W. A. Simson**, H. Feussner, S.T. Kim, N. Navab. *"TeCNO: Surgical Phase Recognition with Multi-stage Temporal Convolutional Networks."* International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Lima, 2020

5. H. Hase*, M.F. Azampour*, M. Tirindelli, M. Paschali, **W. A. Simson**, E. Fatemizadeh, N. Navab. *"Ultrasound-Guided Robotic Navigation with Deep Reinforcement Learning."*

IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, 2020 (Equal Contribution)

6. M. Paschali, **W. A. Simson**, A. Guha Roy, R. Göbl, C. Wachinger, N. Navab. *"Manifold Exploring Data Augmentation with Geometric Transformations for Increased Performance and Robustness."* International Conference on Information Processing in Medical Imaging (IPMI), Hong Kong, 2019

7. M. Paschali*, M.F. Naeem*, **W. A. Simson**, K. Steiger, M. Mollenhauer, N. Navab. *"Deep Learning Under the Microscope: Improving the Interpretability of Medical Imaging Neural Networks."*, arXiv preprint arXiv:1904.03127, 2019 (Equal Contribution)

8. J. Esteban, **W. A. Simson**, S. R. Witzig, A. Rienmüller, S. Virga, B. Frisch, O. Zettinig, D. Sakara, Y. Ryang, N. Navab, C. Hennersperger. *"Robotic ultrasound-guided facet joint insertion."* International journal of computer assisted radiology and surgery (IJCARS), 2018

# B

# Abstracts of Publications not Discussed in this Thesis

## Deep Learning Beamforming for Sub-sampled Ultrasound data

**W. A. Simson**, M. Paschali, N. Navab, G. Zahnd. IEEE International Ultrasonics Symposium (IUS), Kobe, 2018



In medical imaging tasks, such as cardiac imaging, ultrasound acquisition time is crucial, however traditional high-quality beamforming techniques are computationally expensive and their performance is hindered by sub-sampled data. To this end, we propose DeepFormer, a method to reconstruct high quality ultrasound images in real-time on sub-sampled raw data by performing an end-to-end deep learning-based reconstruction. Results on an in vivo dataset of 19 participants show that DeepFormer offers promising advantages over traditional processing of sub-sampled raw-ultrasound data and produces reconstructions that are both qualitatively and visually equivalent to fully-sampled DeepFormed images.

## End-to-end Learning-based Ultrasound Reconstruction

**W. A. Simson**, R. Göbl, M. Paschali, M. Krönke, K. Scheidhauer, W. Weber, N. Navab. arXiv preprint arXiv/1904.04696, 2019

Ultrasound imaging is caught between the quest for the highest image quality, and the necessity for clinical usability. Our contribution is two-fold: First, we propose a novel fully convolutional neural network for ultrasound reconstruction. Second, a custom loss function tailored to the modality is employed for end-to-end training of the network. We demonstrate that training a network to map time-delayed raw data to a minimum variance ground truth offers performance increases in a clinical environment. In doing so, a path is explored towards improved clinically viable ultrasound reconstruction. The proposed method displays both promising image reconstruction quality and acquisition frequency when integrated for live ultrasound scanning. A clinical evaluation is conducted to verify the diagnostic usefulness of the proposed method in a clinical setting.

## Acoustic Shadowing Aware Robotic Ultrasound: Lighting up the Dark

V. Sutedjo, M. Tirindelli, C. Eilers, **W. Simson**, B. Busam, N. Navab. IEEE Robotics and Automation Letters, 2022



Medical Ultrasound (US), despite its wide use, is characterized by artifacts and operator dependency. Those attributes hinder the gathering and utilization of US datasets for the training of Deep Neural Networks used for Computer-Assisted Intervention Systems. Data augmentation is commonly used to enhance model generalization and performance. However, common data augmentation techniques, such as affine transformations do not align with the physics of US and, when used carelessly can lead to unrealistic US images. To this end, we propose a set of physics-inspired transformations, including deformation, reverb and Signal-to-Noise Ratio, that we apply on US B-mode images for data augmentation. We evaluate our method on a new spine US dataset for the tasks of bone segmentation and classification.

## Rethinking Ultrasound Augmentation: A Physics-Inspired Approach

M. Tirindelli*, C. Eilers*, **W. A. Simson**, M. Paschali, M.F. Azampour, N. Navab. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Strasbourg, 2021 (Equal Contribution)



Medical Ultrasound (US), despite its wide use, is characterized by artifacts and operator dependency. Those attributes hinder the gathering and utilization of US datasets for the training of Deep Neural Networks used for Computer-Assisted Intervention Systems. Data augmentation is commonly used to enhance model generalization and performance. However, common data augmentation techniques, such as affine transformations do not align with the physics of US and, when used carelessly can lead to unrealistic US images. To this end, we propose a set of physics-inspired transformations, including deformation, reverb and Signal-to-Noise Ratio, that we apply on US B-mode images for data augmentation. We evaluate our method on a new spine US dataset for the tasks of bone segmentation and classification.

## Automatic Normal Positioning of Robotic Ultrasound Probe based only on Confidence Map Optimization and Force Measurement

Z. Jiang, M. Grimm, M. Zhou, J. Esteban, **W. A. Simson**, G. Zahnd, N. Navab. IEEE Robotics and Automation Letters (RAL), 2020



Acquiring good image quality is one of the main challenges for fully-automatic robot-assisted ultrasound systems (RUSS). The presented method aims at overcoming this challenge for orthopaedic applications by optimizing the orientation of the robotic ultrasound (US) probe, i.e. aligning the central axis of the US probe to the tissue's surface normal at the point of contact in order to improve sound propagation within the tissue. We first optimize the in-plane orientation of the probe by analyzing the confidence map of the US image. We then carry out a fan motion and analyze the resulting forces estimated from joint torques to align the central axis of the probe to the normal within the plane orthogonal to the initial image plane. This results in the final 3D alignment of the probe's main axis with the normal to the anatomical surface at the point of contact without using external sensors for surface reconstruction or localizing the point of contact in an anatomical atlas. The algorithm is evaluated both on a phantom and on human tissues (forearm, upper arm and lower back). The mean absolute angular difference (±STD) between true and estimated normal on stationary phantom, forearm, upper arm and lower back was 3.1 ± 1.0°, 3.7 ± 1.7°, 5.3 ± 1.3° and 6.9 ± 3.5°, respectively. In comparison, six human operators obtained errors of 3.2 ± 1.7° on the phantom. Hence the method is able to automatically position the probe normal to the scanned tissue at the point of contact and thus improve the quality of automatically acquired ultrasound images.

## TeCNO: Surgical Phase Recognition with Multi-Stage Temporal Convolutional Networks

T. Czempiel, M. Paschali, M. Keicher, **W. A. Simson**, H. Feussner, S.T. Kim, N. Navab. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Lima, 2020



Automatic surgical phase recognition is a challenging and crucial task with the potential to improve patient safety and become an integral part of intra-operative decision-support systems. In this paper, we propose, for the first time in workflow analysis, a Multi-Stage Temporal Convolutional Network (MS-TCN) that performs hierarchical prediction refinement for surgical phase recognition. Causal, dilated convolutions allow for a large receptive field and online inference with smooth predictions even during ambiguous transitions. Our method

is thoroughly evaluated on two datasets of laparoscopic cholecystectomy videos with and without the use of additional surgical tool information. Outperforming various state-of-the-art LSTM approaches, we verify the suitability of the proposed causal MS-TCN for surgical phase recognition.

## Ultrasound-Guided Robotic Navigation with Deep Reinforcement Learning

H. Hase*, M.F. Azampour*, M. Tirindelli, M. Paschali, **W. A. Simson**, E. Fatemizadeh, N. Navab. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, 2020 (Equal Contribution)



In this paper we introduce the first reinforcement learning (RL) based robotic navigation method which utilizes ultrasound (US) images as an input. Our approach combines state-of-the-art RL techniques, specifically deep Q-networks (DQN) with memory buffers and a binary classifier for deciding when to terminate the task. Our method is trained and evaluated on an in-house collected data-set of 34 volunteers and when compared to pure RL and supervised learning (SL) techniques, it performs substantially better, which highlights the suitability of RL navigation for US-guided procedures. When testing our proposed model, we obtained a 82.91% chance of navigating correctly to the sacrum from 165 different starting positions on 5 different unseen simulated environments.

## Manifold Exploring Data Augmentation with Geometric Transformations for Increased Performance and Robustness

M. Paschali, **W. A. Simson**, A. Guha Roy, R. Göbl, C. Wachinger, N. Navab. International Conference on Information Processing in Medical Imaging (IPMI), Hong Kong, 2019



In this paper we propose a novel augmentation technique that improves not only the performance of deep neural networks on clean test data, but also significantly increases their robustness to random transformations, both affine and projective. Inspired by ManiFool, the augmentation is performed by a line-search manifold-exploration method that learns affine geometric transformations that lead to the misclassification on an image, while ensuring that it remains on the same manifold as the training data.

This augmentation method populates any training dataset with images that lie on the border of the manifolds between two-classes and maximizes the variance the network is exposed to during training. Our method was thoroughly evaluated on the challenging tasks of fine-grained skin lesion classification from limited data, and breast tumor classification of mammograms.

Compared with traditional augmentation methods, and with images synthesized by Generative Adversarial Networks our method not only achieves state-of-the-art performance but also significantly improves the network's robustness.

## Deep Learning Under the Microscope: Improving the Interpretability of Medical Imaging Neural Networks

M. Paschali*, M.F. Naeem*, **W. A. Simson**, K. Steiger, M. Mollenhauer, N. Navab. arXiv preprint arXiv:1904.03127, 2019 (Equal Contribution)

In this paper, we propose a novel interpretation method tailored to histological Whole Slide Image (WSI) processing. A Deep Neural Network (DNN), inspired by Bag-of-Features models is equipped with a Multiple Instance Learning (MIL) branch and trained with weak supervision for WSI classification. MIL avoids label ambiguity and enhances our model's expressive power without guiding its attention. We utilize a fine-grained logit heatmap of the models activations to interpret its decision-making process. The proposed method is quantitatively and qualitatively evaluated on two challenging histology datasets, outperforming a variety of baselines. In addition, two expert pathologists were consulted regarding the interpretability provided by our method and acknowledged its potential for integration into several clinical applications.

## Robotic Ultrasound-guided Facet Joint Insertion

J. Esteban, **W. A. Simson**, S. R. Witzig, A. Rienmüller, S. Virga, B. Frisch, O. Zettinig, D. Sakara, Y. Ryang, N. Navab, C. Hennersperger. International journal of computer assisted radiology and surgery (IJCARS), 2018

Abstract used with permission from Springer Nature Customer Service Centre GmbH with license number: 5264940981782

**Purpose** Facet joint insertion is a common treatment of chronic pain in the back and spine. This procedure is often performed under fluoroscopic guidance, where the staff's repetitive radiation exposure remains an unsolved problem. Robotic ultrasound (rUS) has the potential to reduce or even eliminate the use of radiation by using ultrasound with a robotic-guided needle insertion. This work presents first clinical data of rUS-based needle insertions extending previous work of our group.

**Methods** Our system implements an automatic US acquisition protocol combined with a calibrated needle targeting system. This approach assists the physician by positioning the needle holder on a trajectory selected in a 3D US volume of the spine.

**Results** By the time of submission, nine facets were treated with our approach as first data from an ongoing clinical study. The insertion success rate was shown to be comparable to current clinical practice. Furthermore, US imaging offers additional anatomical context for needle trajectory planning.

**Conclusion** This work shows first clinical data for robotic ultrasound-assisted facet joint insertion as a promising solution that can easily be incorporated into the clinical workflow. Presented results show the clinical value of such a system.

The cover has been designed using resources from Flaticon.com

# Bibliography


[1] R. Ali and J. J. Dahl. “Travel-time tomography for local sound speed reconstruction using average sound speeds”. In: *2019 IEEE International Ultrasonics Symposium (IUS)*. IEEE. 2019, pp. 2007–2010 (cit. on p. 57).

[2] M. E. Anderson and G. E. Trahey. “The direct estimation of sound speed using pulse–echo ultrasound”. In: *The Journal of the Acoustical Society of America* 104.5 (1998), pp. 3099–3106 (cit. on pp. 55, 56).

[3] A. Antoniou, A. Storkey, and H. Edwards. “Data augmentation generative adversarial networks”. In: *arXiv preprint arXiv:1711.04340* (2017) (cit. on p. 30).

[4] A. Baghani, S. Salcudean, M. Honarvar, R. S. Sahebjavaher, R. Rohling, and R. Sinkus. “Travelling wave expansion: a model fitting approach to the inverse problem of elasticity reconstruction”. In: *IEEE Transactions on Medical Imaging* 30.8 (2011), pp. 1555–1565 (cit. on p. 53).

[5] J. Bamber. “Ultrasonic propagation properties of the breast”. In: *Ultrasonic Examination of the breast* (1983), pp. 37–44 (cit. on pp. 48, 77).

[6] J. Bamber and C. Hill. *Physical principles of medical ultrasonics*. 2004 (cit. on p. 38).

[7] M. A. L. Bell, J. Huang, D. Hyun, Y. C. Eldar, R. van Sloun, and M. Mischi. “Challenge on ultrasound beamforming with deep learning (CUBDL)”. In: *2020 IEEE International Ultrasonics Symposium (IUS)*. IEEE. 2020, pp. 1–5 (cit. on p. 81).

[8] J.-P. Berenger. “A perfectly matched layer for the absorption of electromagnetic waves”. In: *Journal of computational physics* 114.2 (1994), pp. 185–200 (cit. on p. 43).

[9] J.-P. Berenger. “Three-dimensional perfectly matched layer for the absorption of electromagnetic waves”. In: *Journal of computational physics* 127.2 (1996), pp. 363–379 (cit. on p. 43).

[10] O. Bogdan, V. Eckstein, F. Rameau, and J.-C. Bazin. “DeepCalib: a deep learning approach for automatic intrinsic calibration of wide field-of-view cameras”. In: *Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production*. 2018 (cit. on p. 87).

[11] N. N. Bojarski. “The k-space formulation of the scattering problem in the time domain”. In: *The Journal of the Acoustical Society of America* 72.2 (1982), pp. 570–584 (cit. on pp. 36, 40, 41).

[12] N. N. Bojarski. “The k-space formulation of the scattering problem in the time domain: An improved single propagator formulation”. In: *The Journal of the Acoustical Society of America* 77.3 (1985), pp. 826–831 (cit. on pp. 36, 40, 41).

[13] L. Bottou and O. Bousquet. “The tradeoffs of large scale learning”. In: *Advances in neural information processing systems* 20 (2007) (cit. on p. 28).

[14] J. P. Boyd. *Chebyshev and Fourier spectral methods*. Courier Corporation, 2001 (cit. on p. 41).

[15] S. Boyd, S. P. Boyd, and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004 (cit. on p. 32).

[16] L. L. Brickson, D. Hyun, M. Jakovljevic, and J. J. Dahl. “Reverberation Noise Suppression in Ultrasound Channel Signals Using a 3D Fully Convolutional Neural Network”. In: *IEEE Transactions on Medical Imaging* 40.4 (2021), pp. 1184–1195 (cit. on pp. 30, 31).

[17] H. Bruns. *Das eikonal*. Vol. 35. S. Hirzel, 1895 (cit. on p. 57).

[18] M. J. Buckingham. “Theory of acoustic attenuation, dispersion, and pulse propagation in unconsolidated granular materials including marine sediments”. In: *The journal of the acoustical society of America* 102.5 (1997), pp. 2579–2596 (cit. on p. 38).

[19] J. T. Burge. *A Basic Introduction To A Basic Introduction To Neural NetworksNeural Networks* (cit. on p. 27).

[20] D. Bychkov, N. Linder, R. Turkki, et al. “Deep learning based tissue analysis predicts outcome in colorectal cancer”. In: *Scientific reports* 8.1 (2018), pp. 1–11 (cit. on p. 30).

[21] S. Campbell. “A short history of sonography in obstetrics and gynaecology”. In: *Facts, views & vision in ObGyn* 5.3 (2013), p. 213 (cit. on p. 4).

[22] J. Capon. “High-resolution frequency-wavenumber spectrum analysis”. In: *Proceedings of the IEEE* 57.8 (1969), pp. 1408–1418 (cit. on p. 32).

[23] M. Caputo. “Linear models of dissipation whose Q is almost frequency independent—II”. In: *Geophysical Journal International* 13.5 (1967), pp. 529–539 (cit. on p. 42).

[24] A. Carovac, F. Smajlovic, and D. Junuzovic. “Application of ultrasound in medicine”. In: *Acta Informatica Medica* 19.3 (2011), p. 168 (cit. on p. 20).

[25] W. Chen and S. Holm. “Physical interpretation of fractional diffusion-wave equation via lossy media obeying frequency power law”. In: *arXiv preprint math-ph/0303040* (2003) (cit. on p. 42).

[26] W. Chen and S. Holm. “Fractional Laplacian time-space models for linear and nonlinear lossy media exhibiting arbitrary frequency power-law dependency”. In: *The Journal of the Acoustical Society of America* 115.4 (2004), pp. 1424–1430 (cit. on p. 42).

[27] S. Cheung and R. Rohling. “Enhancement of needle visibility in ultrasound-guided percutaneous procedures”. In: *Ultrasound in medicine & biology* 30.5 (2004), pp. 617–624 (cit. on p. 4).

[28] Clawpack Development Team. *Clawpack software*. Version 5.7.1. 2020 (cit. on p. 35).

[29] R. S. Cobbold. *Foundations of biomedical ultrasound*. Oxford university press, 2006 (cit. on pp. 20, 21).

[30] B. T. Cox, S. Kara, S. R. Arridge, and P. C. Beard. “k-space propagation models for acoustically heterogeneous media: Application to biomedical photoacoustics”. In: *The Journal of the Acoustical Society of America* 121.6 (2007), pp. 3453–3464 (cit. on p. 41).

[31] G. Cybenko. “Approximation by superpositions of a sigmoidal function”. In: *Mathematics of control, signals and systems* 2.4 (1989), pp. 303–314 (cit. on p. 27).

[32] T. Czempiel, M. Paschali, M. Keicher, et al. “Tecno: Surgical phase recognition with multi-stage temporal convolutional networks”. In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2020, pp. 343–352 (cit. on p. 30).

[33] J. Dahl. *Lecture notes on Advanced Ultrasound Imaging*. Feb. 2019 (cit. on pp. 13, 14, 20–22, 78).

[34] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. “Arcface: Additive angular margin loss for deep face recognition”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 4690–4699 (cit. on p. 30).

[35] L. Deng, G. Hinton, and B. Kingsbury. “New types of deep neural network learning for speech recognition and related applications: An overview”. In: *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE. 2013, pp. 8599–8603 (cit. on p. 27).

[36] F. Duck. “Tissue non-linearity”. In: *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 224.2 (2010), pp. 155–170 (cit. on p. 12).

[37] J. Esteban, W. Simson, S. Requena Witzig, et al. “Robotic ultrasound-guided facet joint insertion”. In: *International journal of computer assisted radiology and surgery* 13.6 (2018), pp. 895–904 (cit. on p. 4).

[38] J. M. Facil, B. Ummenhofer, H. Zhou, L. Montesano, T. Brox, and J. Civera. “CAM-Convs: Camera-Aware Multi-Scale Convolutions for Single-View Depth”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019 (cit. on p. 87).

[39] J. Fan, J. Wang, Z. Chen, C. Hu, Z. Zhang, and W. Hu. “Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique”. In: *Medical physics* 46.1 (2019), pp. 370–381 (cit. on p. 30).

[40] M. Feigin, D. Freedman, and B. W. Anthony. “A deep learning framework for single-sided sound speed inversion in medical ultrasound”. In: *IEEE Transactions on Biomedical Engineering* 67.4 (2019), pp. 1142–1151 (cit. on pp. 30, 59, 61, 62, 68).

[41] B. Fornberg. “Generation of finite difference formulas on arbitrarily spaced grids”. In: *Mathematics of computation* 51.184 (1988), pp. 699–706 (cit. on p. 41).

[42] B. Fornberg. “High-order finite differences and the pseudospectral method on staggered grids”. In: *SIAM Journal on Numerical Analysis* 27.4 (1990), pp. 904–918 (cit. on p. 41).

[43] B. Fornberg. “The pseudospectral method: Comparisons with finite differences for the elastic wave equation”. In: *Geophysics* 52.4 (1987), pp. 483–501 (cit. on p. 41).

[44] D. Foster, M. Arditi, F. Foster, M. Patterson, and J. Hunt. “Computer simulations of speckle in B-scan images”. In: *Ultrasonic imaging* 5.4 (1983), pp. 308–330 (cit. on p. 21).

[45] S. L. Garrett. “Nonlinear Acoustics”. In: *Understanding Acoustics*. Springer, 2020, pp. 701–753 (cit. on p. 39).

[46] J.-L. Gennisson, T. Deffieux, M. Fink, and M. Tanter. “Ultrasound elastography: principles and techniques”. In: *Diagnostic and interventional imaging* 94.5 (2013), pp. 487–495 (cit. on p. 54).

[47] X. Glorot, A. Bordes, and Y. Bengio. “Deep sparse rectifier neural networks”. In: *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings. 2011, pp. 315–323 (cit. on p. 27).

[48] O. Godin. “An effective quiescent medium for sound propagating through an inhomogeneous, moving fluid”. In: *The Journal of the Acoustical Society of America* 112 (Nov. 2002), pp. 1269–75 (cit. on p. 37).

[49] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016 (cit. on pp. 28, 66, 78).

[50] J. W. Goodman. “Statistical Optics (Wiley Classics Library)”. In: (2000) (cit. on p. 14).

[51] D. Gottlieb and J. S. Hesthaven. “Spectral methods for hyperbolic problems”. In: *Journal of Computational and Applied Mathematics* 128.1-2 (2001), pp. 83–131 (cit. on p. 41).

[52] D. Gottlieb and E. Tadmor. “The CFL condition for spectral approximations to hyperbolic initial-boundary value problems”. In: *Mathematics of Computation* 56.194 (1991), pp. 565–588 (cit. on p. 41).

[53] S. Goudarzi, A. Asif, and H. Rivaz. “Fast multi-focus ultrasound image recovery using generative adversarial networks”. In: *IEEE Transactions on Computational Imaging* 6 (2020), pp. 1272–1284 (cit. on p. 31).

[54] S. Goudarzi, A. Asif, and H. Rivaz. “Multi-focus ultrasound imaging using generative adversarial networks”. In: *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*. IEEE. 2019, pp. 1118–1121 (cit. on p. 30).

[55] J. F. Greenleaf, M. Fatemi, and M. Insana. “Selected methods for imaging elastic properties of biological tissues”. In: *Annual review of biomedical engineering* 5.1 (2003), pp. 57–78 (cit. on p. 54).

[56] H. Greenspan, G. Oz, N. Kiryati, and S. Peled. “MRI inter-slice reconstruction using super-resolution”. In: *Magnetic resonance imaging* 20.5 (2002), pp. 437–446 (cit. on p. 30).

[57] M. F. Hamilton and D. T. Blackstock. “On the coefficient of nonlinearity $\beta$ in nonlinear acoustics”. In: *The Journal of the Acoustical Society of America* 83.1 (1988), pp. 74–77 (cit. on p. 39).

[58] H. Hase, M. F. Azampour, M. Tirindelli, et al. “Ultrasound-guided robotic navigation with deep reinforcement learning”. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020, pp. 5534–5541 (cit. on pp. 4, 30).

[59] P. Hasgall, D. Gennaro, C. Baumgartner, E. Neufeld, et al. *IT’IS Database for thermal and electromagnetic parameters of biological tissues, Version 4.0*. 2018 (cit. on p. 77).

[60] M. H. Hassoun et al. *Fundamentals of artificial neural networks*. MIT press, 1995 (cit. on p. 27).

[61] S. Haykin. “Neural networks: a comprehensive foundation prentice-hall upper saddle river”. In: *NJ MATH Google Scholar* (1999) (cit. on p. 43).

[62] K. He, X. Zhang, S. Ren, and J. Sun. “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1026–1034 (cit. on p. 65).

[63] K. Hornik. “Approximation capabilities of multilayer feedforward networks”. In: *Neural networks* 4.2 (1991), pp. 251–257 (cit. on p. 27).

[64] P. R. Hoskins, K. Martin, and A. Thrush. *Diagnostic ultrasound: physics and equipment*. CRC Press, 2019 (cit. on pp. 8–11, 17–22, 24–26, 78).

[65] D. H. Howry. “The ultrasonic visualization of soft tissue structures and disease processes”. In: *JOURNAL OF LABORATORY AND CLINICAL MEDICINE*. Vol. 40. 5. MOSBY-YEAR BOOK INC 11830 WESTLINE INDUSTRIAL DR, ST LOUIS, MO 63146-3318. 1952, pp. 812–813 (cit. on p. 4).

[66] S.-Y. Huang, J. M. Boone, K. Yang, A. L. Kwan, and N. J. Packard. “The effect of skin thickness determined using breast CT on mammographic dosimetry”. In: *Medical physics* 35.4 (2008), pp. 1199–1206 (cit. on p. 48).

[67] V. F. Humphrey. “Nonlinear propagation in ultrasonic fields: measurements, modelling and harmonic imaging”. In: *Ultrasonics* 38.1-8 (2000), pp. 267–272 (cit. on p. 12).

[68] C. Huygens. *Treatise on Light*. Ed. by S. P. Thompson (cit. on p. 22).

[69] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl. “Beamforming and speckle reduction using neural networks”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 66.5 (2019), pp. 898–910 (cit. on pp. 30, 31, 63, 64).

[70] D. Hyun, A. Wiacek, S. Goudarzi, et al. “Deep Learning for Ultrasound Image Formation: CUBDL Evaluation Framework & Open Datasets”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* (2021) (cit. on p. 30).

[71] S. Ioffe and C. Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. In: *International conference on machine learning*. PMLR. 2015, pp. 448–456 (cit. on pp. 59, 64).

[72] M. Jaeger and M. Frenz. “Towards clinical computed ultrasound tomography in echo-mode: Dynamic range artefact reduction”. In: *Ultrasonics* 62 (2015), pp. 299–304 (cit. on p. 58).

[73] M. Jakovljevic, S. Hsieh, R. Ali, G. Chau Loo Kung, D. Hyun, and J. J. Dahl. “Local speed of sound estimation in tissue using pulse-echo ultrasound: Model-based approach”. In: *The Journal of the Acoustical Society of America* 144.1 (2018), pp. 254–266 (cit. on pp. 56, 80).

[74] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun. “What is the best multi-stage architecture for object recognition?” In: *2009 IEEE 12th international conference on computer vision*. IEEE. 2009, pp. 2146–2153 (cit. on p. 27).

[75] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio. “The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017, pp. 11–19 (cit. on p. 65).

[76] J. A. Jensen. “A model for the propagation and scattering of ultrasound in tissue”. In: *The Journal of the Acoustical Society of America* 89.1 (1991), pp. 182–190 (cit. on p. 35).

[77] Z. Jiang, M. Grimm, M. Zhou, et al. “Automatic normal positioning of robotic ultrasound probe based only on confidence map optimization and force measurement”. In: *IEEE Robotics and Automation Letters* 5.2 (2020), pp. 1342–1349 (cit. on p. 4).

[78] F. K. Jush, P. M. Dueppenbecker, and A. Maier. “Data-Driven Speed-of-Sound Reconstruction for Medical Ultrasound: Impacts of Training Data Format and Imperfections on Convergence”. In: *Annual Conference on Medical Image Understanding and Analysis*. Springer. 2021, pp. 140–150 (cit. on p. 61).

[79] Y. M. Kadah, A. A. Farag, J. M. Zurada, A. M. Badawi, and A.-B. Youssef. “Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images”. In: *IEEE transactions on Medical Imaging* 15.4 (1996), pp. 466–478 (cit. on p. 13).

[80] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001 (cit. on p. 54).

[81] E. Kang, J. Min, and J. C. Ye. “A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction”. In: *Medical physics* 44.10 (2017), e360–e375 (cit. on p. 30).

[82] J. F. Kelly, R. J. McGough, and M. M. Meerschaert. “Analytical time-domain Green’s functions for power-law media”. In: *The Journal of the Acoustical Society of America* 124.5 (2008), pp. 2861–2872 (cit. on p. 42).

[83] S. Khan, J. Huh, and J. C. Ye. “Deep learning-based universal beamformer for ultrasound imaging”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2019, pp. 619–627 (cit. on p. 31).

[84] S. Khan, J. Huh, and J. C. Ye. “Universal deep beamformer for variable rate ultrasound imaging”. In: *arXiv preprint arXiv:1901.01706* (2019) (cit. on p. 30).

[85] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014) (cit. on pp. 59, 69).

[86] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders. *Fundamentals of acoustics*. John wiley & sons, 2000 (cit. on p. 8).

[87] A. Krizhevsky, I. Sutskever, and G. E. Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems* 25 (2012) (cit. on p. 30).

[88] I. Kuo, B. Hete, and K. Shung. “A novel method for the measurement of acoustic speed”. In: *The Journal of the Acoustical Society of America* 88.4 (1990), pp. 1679–1682 (cit. on p. 70).

[89] Y. LeCun, Y. Bengio, et al. “Convolutional networks for images, speech, and time series”. In: *The handbook of brain theory and neural networks* 3361.10 (1995), p. 1995 (cit. on pp. 29, 59).

[90] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning”. In: *nature* 521.7553 (2015), pp. 436–444 (cit. on p. 29).

[91] Y. LeCun, B. Boser, J. S. Denker, et al. “Backpropagation applied to handwritten zip code recognition”. In: *Neural computation* 1.4 (1989), pp. 541–551 (cit. on p. 29).

[92] M. A. Lediju, G. E. Trahey, B. C. Byram, and J. J. Dahl. “Short-lag spatial coherence of backscattered echoes: Imaging characteristics”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 58.7 (2011), pp. 1377–1388 (cit. on p. 14).

[93] J. Lee, Y. Yoo, C. Yoon, and T.-k. Song. “A computationally efficient mean sound speed estimation method based on an evaluation of focusing quality for medical ultrasound imaging”. In: *Electronics* 8.11 (2019), p. 1368 (cit. on p. 64).

[94] M. Levitt and A. Warshel. “Computer simulation of protein folding”. In: *Nature* 253.5494 (1975), pp. 694–698 (cit. on p. 35).

[95] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. A. Jensen, and O. Bernard. “Plane-wave imaging challenge in medical ultrasound”. In: *2016 IEEE International ultrasonics symposium (IUS)*. IEEE. 2016, pp. 1–4 (cit. on p. 81).

[96] M. Liebler, S. Ginter, T. Dreyer, and R. E. Riedlinger. “Full wave modeling of therapeutic ultrasound: Efficient time-domain implementation of the frequency power-law attenuation”. In: *The Journal of the Acoustical Society of America* 116.5 (2004), pp. 2742–2750 (cit. on p. 42).

[97] M. Lienen and S. Günnemann. “Learning the Dynamics of Physical Systems from Sparse Observations with Finite Element Networks”. In: *International Conference on Learning Representations*. 2021 (cit. on p. 87).

[98] J. Long, E. Shelhamer, and T. Darrell. “Fully convolutional networks for semantic segmentation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440 (cit. on p. 30).

[99] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang. “The expressive power of neural networks: A view from the width”. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017, pp. 6232–6240 (cit. on p. 27).

[100] A. C. Luchies and B. C. Byram. “Deep neural networks for ultrasound beamforming”. In: *IEEE transactions on medical imaging* 37.9 (2018), pp. 2010–2021 (cit. on pp. 30, 31).

[101] B. Luijten, R. Cohen, F. J. de Bruijn, et al. “Adaptive ultrasound beamforming using deep learning”. In: *IEEE Transactions on Medical Imaging* 39.12 (2020), pp. 3967–3978 (cit. on p. 30).

[102] B. Luijten, R. Cohen, F. J. de Bruijn, et al. “Deep learning for fast adaptive beamforming”. In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 1333–1337 (cit. on p. 30).

[103] B. Luijten, R. Cohen, F. J. de Bruijn, et al. “Deep learning for fast adaptive beamforming”. In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 1333–1337 (cit. on p. 32).

[104] E. Macé, G. Montaldo, I. Cohen, M. Baulac, M. Fink, and M. Tanter. “Functional ultrasound imaging of the brain”. In: *Nature methods* 8.8 (2011), pp. 662–664 (cit. on pp. 53, 54).

[105] R. Mallart and M. Fink. “Adaptive focusing in scattering media through sound-speed inhomogeneities: The van Cittert Zernike approach and focusing criterion”. In: *The Journal of the Acoustical Society of America* 96.6 (1994), pp. 3721–3732 (cit. on pp. 13, 57).

[106] R. Mallart and M. Fink. “The van Cittert–Zernike theorem in pulse echo measurements”. In: *The Journal of the Acoustical Society of America* 90.5 (1991), pp. 2718–2727 (cit. on p. 14).

[107] W. Marczak. “Water as a standard in the measurements of speed of sound in liquids”. In: *the Journal of the Acoustical Society of America* 102.5 (1997), pp. 2776–2779 (cit. on p. 70).

[108] T. D. Mast, L. P. Souriau, D.-L. Liu, M. Tabei, A. I. Nachman, and R. C. Waag. “A k-space method for large-scale models of wave propagation in tissue”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 48.2 (2001), pp. 341–354 (cit. on pp. 36, 40, 41).

[109] *Measuring the bandwidth and fractional bandwidth of an ultrasound array*. July 2020 (cit. on p. 21).

[110] F. Milletari, N. Navab, and S.-A. Ahmadi. “V-net: Fully convolutional neural networks for volumetric medical image segmentation”. In: *2016 fourth international conference on 3D vision (3DV)*. IEEE. 2016, pp. 565–571 (cit. on p. 30).

[111] S. Min, B. Lee, and S. Yoon. “Deep learning in bioinformatics”. In: *Briefings in bioinformatics* 18.5 (2017), pp. 851–869 (cit. on p. 27).

[112] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink. “Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 56.3 (2009), pp. 489–506 (cit. on pp. 23, 68).

[113] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos. “ORB-SLAM: a versatile and accurate monocular SLAM system”. In: *IEEE transactions on robotics* 31.5 (2015), pp. 1147–1163 (cit. on p. 30).

[114] K. P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012 (cit. on p. 78).

[115] A. A. Nair, T. D. Tran, A. Reiter, and M. A. L. Bell. “A generative adversarial neural network for beamforming ultrasound images: Invited presentation”. In: *2019 53rd Annual conference on information sciences and systems (CISS)*. IEEE. 2019, pp. 1–6 (cit. on p. 30).

[116] V. Nair and G. E. Hinton. “Rectified linear units improve restricted boltzmann machines”. In: *Icml*. 2010 (cit. on p. 27).

[117] S. Narayan. “The generalized sigmoid activation function: Competitive supervised learning”. In: *Information sciences* 99.1-2 (1997), pp. 69–82 (cit. on p. 27).

[118] J. Nebeker and T. R. Nelson. “Imaging of sound speed using reflection ultrasound tomography”. In: *Journal of Ultrasound in Medicine* 31.9 (2012), pp. 1389–1404 (cit. on p. 77).

[119] K. Nightingale. “Acoustic radiation force impulse (ARFI) imaging: a review”. In: *Current medical imaging* 7.4 (2011), pp. 328–339 (cit. on p. 54).

[120] L. Nock, G. E. Trahey, and S. W. Smith. “Phase aberration correction in medical ultrasound using speckle brightness as a quality factor”. In: *The Journal of the Acoustical Society of America* 85.5 (1989), pp. 1819–1833 (cit. on p. 70).

[121] M. L. Oelze and J. Mamou. “Review of quantitative ultrasound: Envelope statistics and backscatter coefficient imaging and contributions to diagnostic ultrasound”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 63.2 (2016), pp. 336–351 (cit. on pp. 4, 35).

[122] S. H. Okazawa, R. Ebrahimi, J. Chuang, R. N. Rohling, and S. E. Salcudean. “Methods for segmenting curved needles in ultrasound images”. In: *Medical image analysis* 10.3 (2006), pp. 330–342 (cit. on p. 4).

[123] J. Page and P. Favaros. “Learning to Model and Calibrate Optics Via a Differentiable Wave Optics Simulator”. In: *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2020, pp. 2995–2999 (cit. on p. 87).

[124] B. Pang, T. Zhao, X. Xie, and Y. N. Wu. “Trajectory prediction with latent belief energy-based model”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 11814–11824 (cit. on p. 30).

[125] M. Paschali, S. Conjeti, F. Navarro, and N. Navab. “Generalizability vs. robustness: investigating medical imaging networks using adversarial examples”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 493–501 (cit. on p. 30).

[126] M. Paschali, W. Simson, A. G. Roy, R. Göbl, C. Wachinger, and N. Navab. “Manifold exploring data augmentation with geometric transformations for increased performance and robustness”. In: *International Conference on Information Processing in Medical Imaging*. Springer. 2019, pp. 517–529 (cit. on p. 30).

[127] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035 (cit. on p. 69).

[128] K. Pearson. “The problem of the random walk”. In: *Nature* 72.1867 (1905), pp. 342–342 (cit. on p. 13).

[129] D. Perdios, A. Besson, M. Arditi, and J.-P. Thiran. “A deep learning approach to ultrasound image recovery”. In: *2017 IEEE International Ultrasonics Symposium (IUS)*. Ieee. 2017, pp. 1–4 (cit. on p. 30).

[130] A. D. Pierce and R. T. Beyer. *Acoustics: An introduction to its physical principles and applications. 1989 Edition*. 1990 (cit. on p. 37).

[131] G. Pinton. “Three dimensional full-wave nonlinear acoustic simulations of ultrasound imaging and therapy in the entire human body”. In: *2012 IEEE International Ultrasonics Symposium*. IEEE. 2012, pp. 142–145 (cit. on p. 36).

[132] G. Pinton. “Ultrasound imaging with three dimensional full-wave nonlinear acoustic simulations. Part 2: sources of image degradation in intercostal imaging”. In: *arXiv preprint arXiv:2003.06927* (2020) (cit. on p. 36).

[133] G. F. Pinton, J. Dahl, S. Rosenzweig, and G. E. Trahey. “A heterogeneous nonlinear attenuating full-wave model of ultrasound”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 56.3 (2009), pp. 474–488 (cit. on pp. 36, 40).

[134] I. Podlubny. *Fractional differential equations: an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications*. Elsevier, 1998 (cit. on p. 42).

[135] X. Qu, T. Azuma, H. Lin, et al. “Limb muscle sound speed estimation by ultrasound computed tomography excluding receivers in bone shadow”. In: *Medical Imaging 2017: Ultrasonic Imaging and Tomography*. Vol. 10139. International Society for Optics and Photonics. 2017, 101391B (cit. on p. 54).

[136] G. Rahbar, A. C. Sie, G. C. Hansen, et al. “Benign versus malignant solid breast masses: US differentiation”. In: *Radiology* 213.3 (1999), pp. 889–894 (cit. on p. 26).

[137] M. Raissi, P. Perdikaris, and G. E. Karniadakis. “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations”. In: *Journal of Computational Physics* 378 (2019), pp. 686–707 (cit. on p. 87).

[138] M. Raissi, P. Perdikaris, and G. E. Karniadakis. “Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations”. In: *arXiv preprint arXiv:1711.10561* (2017) (cit. on p. 87).

[139] M. Raissi, P. Perdikaris, and G. E. Karniadakis. “Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations”. In: *arXiv preprint arXiv:1711.10566* (2017) (cit. on p. 87).

[140] R. Ribeiro and J. Sanches. “Fatty liver characterization and classification by ultrasound”. In: *Iberian conference on pattern recognition and image analysis*. Springer. 2009, pp. 354–361 (cit. on p. 13).

[141] D. Robinson, J. Ophir, L. Wilson, and C. Chen. “Pulse-echo ultrasound speed measurements: progress and prospects”. In: *Ultrasound in medicine & biology* 17.6 (1991), pp. 633–646 (cit. on p. 55).

[142] A. Rodriguez-Molares, O. M. H. Rindal, J. D’hooge, S.-E. Måsøy, A. Austeng, and H. Torp. “The generalized contrast-to-noise ratio”. In: *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE. 2018, pp. 1–4 (cit. on p. 25).

[143] O. Ronneberger, P. Fischer, and T. Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241 (cit. on p. 30).

[144] O. Ronneberger, P. Fischer, and T. Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241 (cit. on p. 65).

[145] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. “Learning representations by back-propagating errors”. In: *nature* 323.6088 (1986), pp. 533–536 (cit. on p. 29).

[146] M. Salehi, S.-A. Ahmadi, R. Prevost, N. Navab, and W. Wein. “Patient-specific 3D ultrasound simulation based on convolutional ray-tracing and appearance optimization”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2015, pp. 510–518 (cit. on p. 35).

[147] S. J. Sanabria, E. Ozkan, M. Rominger, and O. Goksel. “Spatial domain reconstruction for imaging speed-of-sound with pulse-echo ultrasound: simulation and in vivo study”. In: *Physics in Medicine & Biology* 63.21 (2018), p. 215015 (cit. on pp. 54, 57).

[148] O. Senouf, S. Vedula, G. Zurakhov, et al. “High frame-rate cardiac ultrasound imaging with deep learning”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 126–134 (cit. on p. 30).

[149] T.-J. Shan and T. Kailath. “Adaptive beamforming for coherent signals and interference”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 33.3 (1985), pp. 527–536 (cit. on p. 32).

[150] Y. Shao, H. Hashemi, P. Gordon, et al. “Breast Cancer Detection using Multimodal Time Series Features from Ultrasound Shear Wave Absolute Vibro-Elastography”. In: *IEEE Journal of Biomedical and Health Informatics* (2021) (cit. on p. 4).

[151] R. E. Sheriff and L. P. Geldart. *Exploration seismology*. Cambridge university press, 1995 (cit. on p. 55).

[152] K. K. Shung and G. A. Thieme. *Ultrasonic scattering in biological tissues*. CRC press, 1992 (cit. on p. 13).

[153] K. Simonyan and A. Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014) (cit. on p. 59).

[154] W. Simson, R. Göbl, M. Paschali, et al. “End-to-end learning-based ultrasound reconstruction”. In: *arXiv preprint arXiv:1904.04696* (2019) (cit. on p. 31).

[155] W. Simson, M. Paschali, N. Navab, and G. Zahnd. “Deep learning beamforming for sub-sampled ultrasound data”. In: *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE. 2018, pp. 1–4 (cit. on pp. 30, 31).

[156] R. Smithuis, L. Wijers, and I. Dennert. *Ultrasound of the Breast*. `https://radiologyassistant.nl/breast/ultrasound/ultrasound-of-the-breast`. Accessed: 2021-08-22 (cit. on pp. 46, 47).

[157] R. Sood, A. F. Rositch, D. Shakoor, et al. “Ultrasound for breast cancer detection globally: a systematic review and meta-analysis”. In: *Journal of global oncology* 5 (2019), pp. 1–17 (cit. on p. 26).

[158] P. Stähli, M. Kuriakose, M. Frenz, and M. Jaeger. “Improved forward model for quantitative pulse-echo speed-of-sound imaging”. In: *Ultrasonics* 108 (2020), p. 106168 (cit. on pp. 58, 64, 80).

[159] K. Sun, B. Xiao, D. Liu, and J. Wang. “Deep High-Resolution Representation Learning for Human Pose Estimation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019 (cit. on p. 30).

[160] J.-F. Synnevag, A. Austeng, and S. Holm. “Benefits of minimum-variance beamforming in medical ultrasound imaging”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 56.9 (2009), pp. 1868–1879 (cit. on p. 32).

[161] T. L. Szabo. “Time domain wave equations for lossy media obeying a frequency power law”. In: *The Journal of the Acoustical Society of America* 96.1 (1994), pp. 491–500 (cit. on p. 42).

[162] M. Tabei, T. D. Mast, and R. C. Waag. “A k-space method for coupled first-order acoustic propagation equations”. In: *The Journal of the Acoustical Society of America* 111.1 (2002), pp. 53–63 (cit. on pp. 40, 41, 43).

[163] R. Taborda and J. Bielak. “Large-scale earthquake simulation: computational seismology and complex engineering systems”. In: *Computing in Science & Engineering* 13.4 (2011), pp. 14–27 (cit. on p. 35).

[164] M. S. Taljanovic, L. H. Gimber, G. W. Becker, et al. “Shear-wave elastography: basic physics and musculoskeletal applications”. In: *Radiographics* 37.3 (2017), pp. 855–870 (cit. on pp. 53, 54).

[165] G. Taraldsen. “A generalized Westervelt equation for nonlinear medical ultrasound”. In: *The Journal of the Acoustical Society of America* 109.4 (2001), pp. 1329–1333 (cit. on pp. 36, 39).

[166] K. J. Taylor, P. N. Burns, and P. N. Well. “Clinical applications of Doppler ultrasound”. In: (1987) (cit. on pp. 53, 54).

[167] J. C. Tillett, M. I. Daoud, J. C. Lacefield, and R. C. Waag. “A k-space method for acoustic propagation using coupled first-order equations in three dimensions”. In: *The Journal of the Acoustical Society of America* 126.3 (2009), pp. 1231–1244 (cit. on p. 41).

[168] L. Tong, H. Gao, H. F. Choi, and J. D’hooge. “Comparison of conventional parallel beamforming with plane wave and diverging wave imaging for cardiac applications: A simulation study”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 59.8 (2012), pp. 1654–1663 (cit. on p. 23).

[169] A. Toshev and C. Szegedy. “Deeppose: Human pose estimation via deep neural networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 1653–1660 (cit. on p. 30).

[170] B. E. Treeby and B. T. Cox. “Modeling power law absorption and dispersion for acoustic propagation using the fractional Laplacian”. In: *The Journal of the Acoustical Society of America* 127.5 (2010), pp. 2741–2748 (cit. on p. 42).

[171] B. E. Treeby and B. T. Cox. “k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields”. In: *Journal of biomedical optics* 15.2 (2010), p. 021314 (cit. on pp. 35, 36, 60, 68, 69).

[172] B. E. Treeby, J. Jaros, A. P. Rendell, and B. Cox. “Modeling nonlinear ultrasound propagation in heterogeneous media with power law absorption using ak-space pseudospectral method”. In: *The Journal of the Acoustical Society of America* 131.6 (2012), pp. 4324–4336 (cit. on p. 36).

[173] L. N. Trefethen. *Spectral methods in MATLAB*. SIAM, 2000 (cit. on p. 41).

[174] E. Turgay, S. Salcudean, and R. Rohling. “Identifying the mechanical properties of tissue by ultrasound strain imaging”. In: *Ultrasound in medicine & biology* 32.2 (2006), pp. 221–235 (cit. on p. 53).

[175] D. Ulyanov, A. Vedaldi, and V. Lempitsky. “Instance normalization: The missing ingredient for fast stylization”. In: *arXiv preprint arXiv:1607.08022* (2016) (cit. on p. 65).

[176] N. Uniyal, H. Eskandari, P. Abolmaesumi, et al. “Ultrasound RF time series for classification of breast lesions”. In: *IEEE transactions on medical imaging* 34.2 (2014), pp. 652–661 (cit. on p. 53).

[177] R. J. Van Sloun, R. Cohen, and Y. C. Eldar. “Deep learning in ultrasound imaging”. In: *Proceedings of the IEEE* 108.1 (2019), pp. 11–29 (cit. on p. 32).

[178] A. Vaswani, S. Bengio, E. Brevdo, et al. “Tensor2tensor for neural machine translation”. In: *arXiv preprint arXiv:1803.07416* (2018) (cit. on p. 27).

[179] S. Vedula, O. Senouf, G. Zurakhov, A. Bronstein, O. Michailovich, and M. Zibulevsky. “Learning beamforming in ultrasound imaging”. In: *arXiv preprint arXiv:1812.08043* (2018) (cit. on p. 30).

[180] F. Vignon and M. R. Burcher. “Capon beamforming in medical ultrasound imaging with focused beams”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 55.3 (2008), pp. 619–628 (cit. on p. 32).

[181] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis. “Deep learning for computer vision: A brief review”. In: *Computational intelligence and neuroscience* 2018 (2018) (cit. on p. 27).

[182] R. F. Wagner, M. F. Insana, and S. W. Smith. “Fundamental correlation lengths of coherent speckle in medical ultrasonic images”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 35.1 (1988), pp. 34–44 (cit. on p. 14).

[183] K. R. Waters, J. Mobley, and J. G. Miller. “Causality-imposed (Kramers-Kronig) relationships between attenuation and dispersion”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 52.5 (2005), pp. 822–823 (cit. on p. 38).

[184] J. Watson, O. M. Aodha, V. Prisacariu, G. Brostow, and M. Firman. “The Temporal Opportunist: Self-Supervised Multi-Frame Monocular Depth”. In: *Computer Vision and Pattern Recognition (CVPR)*. 2021 (cit. on p. 30).

[185] K. A. Wear, R. F. Wagner, D. G. Brown, and M. F. Insana. “Statistical properties of estimates of signal-to-noise ratio and number of scatterers per resolution cell”. In: *The Journal of the Acoustical Society of America* 102.1 (1997), pp. 635–641 (cit. on p. 15).

[186] J. F. Wendt. *Computational fluid dynamics: an introduction*. Springer Science & Business Media, 2008 (cit. on p. 35).

[187] P. J. Westervelt. “Parametric acoustic array”. In: *The Journal of the acoustical society of America* 35.4 (1963), pp. 535–537 (cit. on pp. 36, 39).

[188] M. G. Wismer. “Finite element analysis of broadband acoustic pulses through inhomogenous media with power law attenuation”. In: *The Journal of the Acoustical Society of America* 120.6 (2006), pp. 3493–3502 (cit. on p. 42).

[189] J. Xiao, K. A. Ehinger, A. Oliva, and A. Torralba. “Recognizing scene viewpoint using panoramic place representation”. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. 2012 (cit. on p. 87).

[190] C. Xie, P. Cox, N. Taylor, and S. LaPorte. “Ultrasonography of thyroid nodules: a pictorial review”. In: *Insights into imaging* 7.1 (2016), pp. 77–86 (cit. on p. 26).

[191] G. Yang, S. Yu, H. Dong, et al. “DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction”. In: *IEEE transactions on medical imaging* 37.6 (2017), pp. 1310–1321 (cit. on p. 30).

[192] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao. “Deep learning for single image super-resolution: A brief review”. In: *IEEE Transactions on Multimedia* 21.12 (2019), pp. 3106–3121 (cit. on p. 30).

[193] T. Young, D. Hazarika, S. Poria, and E. Cambria. “Recent trends in deep learning based natural language processing”. In: *ieee Computational intelligenCe magazine* 13.3 (2018), pp. 55–75 (cit. on p. 27).

[194] X. Yuan, D. Borup, J. Wiskin, M. Berggren, and S. A. Johnson. “Simulation of acoustic wave propagation in dispersive media with relaxation losses by using FDTD method with PML absorbing boundary condition”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 46.1 (1999), pp. 14–23 (cit. on p. 43).

[195] J. Zhang, Y. Xie, Y. Xia, and C. Shen. “Attention residual learning for skin lesion classification”. In: *IEEE transactions on medical imaging* 38.9 (2019), pp. 2092–2103 (cit. on p. 30).

[196] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu. “Object detection with deep learning: A review”. In: *IEEE transactions on neural networks and learning systems* 30.11 (2019), pp. 3212–3232 (cit. on p. 30).

[197] O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu. *The finite element method: its basis and fundamentals*. Elsevier, 2005 (cit. on p. 35).

# List of Figures

# List of Tables