Implementation of Shear Wave Elastography for Point-of-Care Ultrasound Imaging

Damian Cacko

A thesis presented for the degree of
Doctor of Philosophy

Thesis supervisor: Prof. Zbigniew Ranachowski
Auxiliary supervisor: Dr Marcin Lewandowski

Abstract

Assessment of tissue stiffness provides a clinical opportunity for screening, early detection, and diagnosis of diseases such as breast cancer or liver fibrosis. Among all ultrasound-based elasticity imaging methods, Shear Wave Elastography (SWE) has emerged as a promising technique that enables quantitative soft tissue stiffness estimation. Novel concepts and methodological approaches in SWE have been the subject of intense research and new system solutions continue to be proposed. However, its practical implementation is complicated and constitutes several engineering challenges, including generation of high-energy excitation (push pulses) and high frame-rate data acquisition. Another key requirement is to provide a sufficient computational performance of the system to process large datasets in an image reconstruction procedure. In particular, in real-time applications, it requires providing both robust computational resources and a smart algorithm implementation strategy. Due to the aforementioned reasons, to date, the SWE has been available only on high-end commercial systems or bulk and expensive research platforms. In particular, SWE mode in the portable ultrasound devices class is not available at all. As a result, the accessibility of this method is limited.

In this thesis, engineering and research efforts were put to address this issue. Firstly, in the theoretical part, the physical principles of the SWE were discussed, and the known methods for ultrasonic tissue elasticity imaging methods were explored. It was followed by a detailed review of the 2-D SWE technique, including acquisition sequences, data processing methods, and considerations on imaging quality and safety. Then, the 2-D SWE method was implemented using a low-cost, portable, and fully configurable research system which was enhanced by the development of a dedicated 256-channel transmit extension module to support push pulses generation. The multiple-stage image reconstruction algorithm was developed that consisted of: RF data filtering, down-conversion, beamforming, shear wave detection, shear wave motion data filtering, shear wave speed estimation, image compounding, and post-processing. Real-time imaging performance was achieved by exploiting massively parallel processing using GPU for computational-intensive algorithm steps, which involved careful problem decomposition. The system was validated in experiments using both homogeneous and heterogeneous regions of an industry-standard elastography phantom. The developed system performance regarding image reconstruction speed was evaluated, and the impact of the most relevant reconstruction parameters on processing time and imaging quality was explored. Since the image reconstruction process was found to bring the biggest contribution to the whole system's power consumption, the obtained results allowed to evaluate the trade-off between power consumption and imaging quality to find a reasonable balance between those factors in a given application for efficient implementation of the 2-D SWE method using a portable scanner.

The overall achieved results have shown that by careful design it is feasible to introduce the real-time 2-D SWE imaging on a portable, cost-optimized, software-oriented ultrasound system.

Abstrakt

Ocena sztywności tkanek jest narzędziem wczesnego wykrywania i diagnozowania chorób, takich jak rak piersi lub zwłóknienie wątroby. Spośród wszystkich metod obrazowania właściwości sprężystych tkanek opartych na ultrasonografii, elastografia fali ścinania (ang.: shear wave elastography, SWE) stała się obiecującą techniką umożliwiającą zarówno jakościowe, jak i ilościowe obrazowanie sztywności obszarów tkankowych. Przez ostatnie dwie dekady nowatorskie koncepcje i podejścia metodologiczne w elastografii fali ścinania były przedmiotem intensywnych badań. Praktyczna implementacja tej techniki jest przedsięwzięciem złożonym ze względu na szereg wyzwań inżynieryjnych, zwłaszcza konieczność transmisji impulsów wysokiej energii (ang.: push pulses) potrzebnych do generacji akustycznej siły promieniowania (and.: acoustic radiation force) i bardzo szybkiej akwizycji danych. Kolejną komplikację stanowi znaczna złożoność obliczeniowa algorytmu rekonstrukcji obrazu oraz potrzeba operacji na dużych zbiorach danych. W szczególności do realizacji obrazowania w czasie rzeczywistym, oprócz zapewnienia wystarczających zasobów obliczeniowych, wymagane jest uważne zaprojektowanie potoku przetwarzania danych ze względu na ograniczenia czasowe. Z powodu wyżej wymienionych przyczyn, do chwili obecnej technika dwuwymiarowej elastografii fali ścinania jest dostępna tylko na zaawansowanych komercyjnych systemach klinicznych lub możliwa do realizacji na nielicznych, kosztownych platformach badawczych. Ponadto technika ta nie została jak dotąd opracowana w klasie przenośnych urządzeń ultrasonograficznych (ang.: point-of-care). W rezultacie dostępność tej metody jest ograniczona.

W niniejszej pracy zaprezentowano wyniki prac projektowo-konstrukcyjnych i badawczych ukierunkowanych na rozwiązanie powyższego problemu. W części teoretycznej pracy omówiono podstawy fizyczne techniki SWE oraz dokonano przeglądu znanych metod ultradźwiękowego obrazowania właściwości sprężystych tkanek. Następnie szczegółowo omówiono poszczególne aspekty techniki dwuwymiarowej elastografii fali ścinania, w tym: stosowane sekwencje nadawczo-odbiorcze, metody przetwarzania danych, czynniki wpływające na jakość obrazowania oraz kwestie bezpieczeństwa badania. W części praktycznej pracy metoda ta została zaimplementowana za pomocą przenośnego, zoptymalizowanego pod kątem niskiego kosztu jednostkowego i w pełni konfigurowalnego systemu obrazowania ultradźwiękowego do zastosowań badawczych. Dla zapewnienia możliwości generacji wysokoenergetycznych impulsów nadawczych system ten rozszerzono o specjalnie zaprojektowany, 256-kanałowy moduł nadawczy. W obrębie warstwy oprogramowania opracowano wieloetapowy algorytm rekonstrukcji obrazu składający się z: filtracji danych RF, demodulacji, kształtowanie wiązki odbiorczej (ang. receive beamforming), detekcji fali ścinania, filtracji danych dotyczących ruchu fal poprzecznych, estymacji przestrzennego rozkładu prędkości grupowej fali ścinania i przetwarzania końcowego. Wydajność obliczeniową pozwalającą zrealizować obrazowanie w czasie rzeczywistym osiągnięto przez dekompozycję problemu obliczeniowego na potrzeby przetwarzania równoległego oraz użycie procesora graficznego (GPU) do akceleracji intensywnych obliczeniowo kroków algorytmu rekonstrukcji obrazu. W części badawczej pracy działanie opracowanego systemu zweryfikowano w szeregu eksperymentów wykorzystujących zarówno jednorodne, jak i niejednorodne obszary fantomu elastograficznego oraz wyznaczono miary jakościowe uzyskiwanych obrazów. Ponadto oceniono wydajność

systemu pod kątem szybkości rekonstrukcji obrazu, a także zbadano wpływ najistotniejszych parametrów rekonstrukcji na czas przetwarzania i jakość obrazowania. Na przykładzie opracowanego systemu pokazano, że obliczeniowy proces rekonstrukcji obrazu wnosi największy wkład w całkowite zużycie energii systemu. Na podstawie uzyskanych wyników rozważono związki między zużyciem energii a jakością obrazowania. Pozwoliło to wysnuć ogólne wnioski pomocne w znalezieniu kompromisu między tymi czynnikami w kontekście efektywnej implementacji metody dwuwymiarowej elastografii fali ścinania w urządzeniach przenośnych.

Ogół wyników uzyskanych w niniejszej pracy pokazał, że implementacja w czasie rzeczywistym złożonej techniki obrazowania, jaką jest ultradźwiękowa elastografia fali ścinania, jest możliwa z użyciem systemu obrazowania klasy point-of-care.

Acknowledgments

I would like to express my sincere thanks and appreciation to all those without whom this work would not have been possible. In particular, I would like to thank the following people:

- Prof. Zbigniew Ranachowski, the supervisor of this thesis — for all the support and for keeping me motivated,

- Dr Marcin Lewandowski — for years of mentoring, strong and consistent guidance, and creating working conditions that allowed me to combine professional work with conducting scientific research resulting in this thesis,

- Piotr Jarosik — for his assistance and expertise in software development,

- Mateusz Walczak and Beata Witek from the us4us team — for their support and advisement in hardware development.

- My other colleagues from the us4us team — Piotr Karwat, Ziemowit Klimonda, and Jakub Rozbicki — for inspirational discussions and sharing their opinions, knowledge, and experience.

Last but not least, I would like to express my deepest gratitude to my wife Marcelina, and my parents for their continuous and unconditional support in life that allowed me to complete this thesis.

Contents

List of Figures

List of Tables

Acronyms

The list of acronyms used in this thesis and their abbreviations:

ADC	Analog-to-Digital Converter
AFE	Analog Front-End
ARF	Acoustic Radiation Force
ARFI	Acoustic Radiation Force Impulse
ASIC	Application-Specific Integrated Circuit
CNR	Contrast to Noise Ratio
CPU	Central Processing Unit
CPWC	Coherent Plane Wave Compounding
CT	Computed Tomography
CUSE	Comb Push Ultrasound Elastography
CW	Continuous Wave
CWS	Crawling Wave Spectroscopy
DOF	Depth of Field
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FOV	Field of View
FPGA	Field-Programmable Gate Array
FRI	Frame Repetition Interval
GPU	Graphics Processing Unit
IC	Integrated Circuit
IP	Intellectual Property
I/Q	In-phase / Quadrature
MI	Mechanical Index

MRI	Magnetic Resonance Imaging
NCC	Normalized Cross-Correlation
POCUS	Point-of-Care Ultrasound
PRF	Pulse Repetition Frequency
PRI	Pulse Repetition Interval
PSF	Point Spread Function
PW	Pulsed Wave
PWI	Plane Wave Imaging
PWM	Pulse Width Modulation
QSPI	Quad Serial Peripheral Interface
RAM	Random Access Memory
RF	Radio Frequency
ROI	Region of Interest
ROE	Region of Excitation
RX	Receive
SDUV	Shearwave Dispersion Ultrasound Vibrometry
SMURF	Spatially Modulated Ultrasound Radiation Force
SNR	Signal to Noise Ratio
SoS	Speed of Sound
SSI	Supersonic Shear Imaging
SWE	Shear Wave Elastograhy
SWEI	Shear Wave Elasticity Imaging
SWS	Shear Wave Speed
TI	Thermal Index
ToF	Time of Flight
TX	Transmit

Chapter 1

Introduction

Manual palpation is a basic diagnostic method that has been known for centuries and is still widely used in general medical practice nowadays. This method dates back to Hippocrates (400 BC), who noted that abdominal swellings that are "*soft, free of pain, and yield when pressed with the finger, are more chronic*" compared with those that "*as are painful, hard, and large, indicate danger of speedy death*" [1]. Probably, a most commonly used palpation routine is looking for tumors or nodules in the breast, which is done as self-examination, intuitively, by pressing tissue by fingers and searching for smaller displacements that indicate stiffer lesions. Indeed, it has been proven by numerous clinical studies, that changes in tissue stiffness can be a result of various diseases, and stiffness itself is an important biomarker of tissue health.

Although palpation is a simple and useful tool for rapid tissue assessment, this method is highly subjective and its usefulness is limited only to superficial structures. During the last three decades, due to recent developments in medical ultrasound, the ancient method of palpation gained new perspectives due to a number of emerging methods for elasticity imaging — ultrasound elastography. Novel methodological approaches have been the subject of intense research and new system solutions continue to be proposed. Areas of applications of elasticity imaging in medical diagnostics are steadily expanding. With the ability to measure the elastic properties of tissues deep within the body by remote means — both quantitatively and qualitatively in some methods — elastography seems to have a great clinical perspective. In addition, it possesses typical advantages of the vast majority of ultrasound techniques — it is fast, portable, low-cost, and easy to use. Thus, elastography gives an opportunity to bring a new value to ultrasound imaging by combining traditional imaging of tissue structure in B-mode imaging and tissue elastic properties mapping.

This introductory chapter provides a brief overview of ultrasound-based elastography techniques - its basic physical principles, milestone developments, and an overview of most common implementation approaches. The key clinical applications are also described. The chapter concludes with the thesis motivation and formulation of the specific aims to be achieved.

1.1 Background

State-of-the-art medical ultrasound imaging stands out from other modalities in providing real-time diagnostic capability at an affordable price while being physically portable.

For more than four decades, it has been routinely used in hospitals and clinics to diagnose a wide variety of diseases. The success of ultrasound as a versatile diagnostic tool can be explained by the fact, that ultrasound imaging possesses some unique characteristics, that are advantageous in comparison to other medical imaging modalities [2]:

- Ultrasound utilizes non-ionizing radiation and is considered safe according to current knowledge.

- Ultrasound imaging devices are less expensive than modalities of similar capabilities like X-ray computed tomography (CT) or magnetic resonance imaging (MRI).

- It is fast and provides real-time imaging *in-situ*.

- It features resolution in a range of sub-millimeter to millimeters, depending on the application, imaging depth, and frequency used.

The advantages listed above that ultrasound can offer make it a diagnostic tool of choice in many medical disciplines as cardiology, obstetrics, gynecology, surgery, pediatrics, radiology, and neurology, to name just a few [2].

The success of medical ultrasound would not be possible without the significant effort of many researchers over more than sixty years. From its beginnings in the late 1940s and early 1950s until today, ultrasound imaging gained tremendous improvement in terms of image quality, robustness, performance, and a number of possible applications. Theoretical and conceptual works in the field of acoustics were accompanied by rapid developments in electronics, that enabled deployment of novel ideas into operational devices. The first commercial B-mode scanners became available in the mid-1960s, including Polish UG-1 [3] — a pioneer system developed in 1964 in the Institute of Fundamental Technological Research of Polish Academy of Science by a team led by Prof. Filipczyński. Those devices including a single transducer mounted on a long movable arm, display, and electronics weighed one ton and were able to obtain a single, biphasic black and white, static B-mode image after a long acquisition lasting minutes [4]. Incorporating the transistors and first integrated circuits revolutionized ultrasound imaging by accelerating the pace of developments and increasing imaging quality and speed. A significant milestone was reached in 1967 by Siemens (Munchen, Germany), which released the first-ever ultrasound scanner allowing real-time image acquisition [5]. In the 1970s availability of electronic digital circuits such as random-access memories (RAM), microprocessors and analog-to-digital converters (ADCs) allowed the implementation of digital scan converters that increased the dynamic range of gray-scale B-mode images [4]. Advances in transducer technology and transferring electronically controlled phased-array technology from radar applications resulted in the first portable ultrasound system capable of performing real-time imaging [6]. By 1980 real-time phased-array imaging systems were commercially available and became forerunners of future cart-based systems of modular architecture. In the 1980s further improvement in imaging quality was driven by rapid inventions in the field of electronics — digital signal processors (DSPs) and application-specific integrated circuits (ASICs). Continuation of this trend from the 1990s till today with developments of more powerful microprocessors (uPs) and high-density field-programmable gate arrays (FPGAs) resulted in improved image quality, higher integration, and faster computation available at lower costs. Imaging systems incorporating these advances evolved into complex modular digital architectures we know today.

(a)   **(b)**   **(c)**

Figure 1.1: Various point-of-care ultrasound systems: (a) GE Logic e (General Electric, Boston, USA) in a form of a compact laptop. (b) Philips Lumify (Philips, Amsterdam, The Netherlands) with frond-end electronics integrated within the probe. (c) Butterfly iQ+ (Butterfly Network, Guilford, USA) — a complete solution in a single probe that can be connected to a mobile phone. It utilizes a CMUT (capacitive micromachined ultrasound transducer) with over 9000 elements [11]

In recent years, the utilization of graphics processing units (GPUs) for massive parallel processing [7] allowed the implementation of the most sophisticated imaging methods and reconstruction algorithms [8]. Software processing accelerated by GPUs and multi-core CPUs was the enabler to full software-based architecture systems [9]. In these types of systems, instead of increasing integrated hardware processing channels, all the data processing is performed in the software by high-performance CPUs and GPUs. This architecture paved a new way to perform high frame-rate (also called *ultrafast*) imaging, in which plane wave insonifications are used rather than building the image in line-by-line acquisitions as in conventional imaging. This approach allows achieving frame rates in the range of several kHz making novel modalities feasible [10]. Two-dimensional (2-D) shear wave elastography is one of them, which will be shown later in this work.

As shown, improvements in electronics keep shaping the future and capabilities of medical ultrasound. The performance growth along with rapid miniaturization of electronics, ASICs in particular, made feasible the implementation of truly portable imaging systems for array transducers with high-quality imaging capabilities. Selected systems from the point-of-care segment are presented in Fig. 1.1. To increase portability, some systems integrate front-end electronics within the probe (Fig. 1.1(b)). Most advanced systems incorporate novel cMUT or pMUT (capacitive and piezoelectric micromachined ultrasound transducer, respectively) arrays, which can be directly integrated with front-end electronics within a single chip containing thousands of transducer elements and processing channels [11]–[13]. An example of such a state-of-art system is shown in Fig. 1.1(c).

As a result, in recent years, the increased availability of low-cost portable ultrasound scanners has opened new application areas for ultrasound, like emergency or primary care, and enhanced the concept of point-of-care ultrasound (POCUS). By point-of-care ultrasound, it is meant "*ultrasound performed at the bedside and interpreted directly by the treating clinician*" [14]. Its highly compact form allows use in any location where medical care could be delivered. Rather than replacing comprehensive ultrasound, POCUS is intended to provide physicians immediate access to clinical imaging

for fast and direct diagnosis. In specific, general practitioners in primary care can potentially benefit from increased accessibility of ultrasound imaging [15], [16]. An overview of possibilities, applications, and future directions of POCUS is available elsewhere [14], [17].

Along with advances in imaging systems, new methods of echo data presentation and signal processing were developed that built inner modalities within ultrasound imaging. Tracking the motion of internal organs like the heart using an M-mode display was first achieved in 1954, and only one year later the possibility of measuring blood flow velocity was shown using continuous wave (CW) Doppler method [4]. Pulsed wave (PW) Doppler has its origins in the mid-1960s but was implemented in imaging systems in the 1980s along with color flow mapping [4]. Further enhancements were possible due to growth in transducer technology. Specialized phased and linear arrays with wider bandwidths were developed for specific clinical applications. From the late 1980s till the early 2000s, there was intensive research done to enable real-time 3-D imaging, that involved 2-D transducer arrays with thousands of elements.

Despite advances in ultrasound imaging system design and the growth of new imaging methods, the feasibility of soft tissue elasticity imaging — ultrasound elastography — was out of focus of ultrasound research until the late 1980s and early 1990s. Before that time research in this field was very limited [4], [18]. However, all the developments that preceded the birth of ultrasound elastography enabled the rapid implementation of selected techniques and their transfer for clinical use. Some of the most complex methods feature high hardware requirements and their implementation became reachable only in recent years using the most advanced hardware platforms.

1.2 Tissue elasticity imaging by ultrasound

Since the early 1990s, there has been an increasing interest observed in the field of tissue elasticity imaging, and a broad portfolio of methods was proposed. Elastography offers imaging of an entirely new subset of tissue properties that cannot be derived with conventional ultrasound techniques. Historical roots of elastography, key developments, and approaches tried through the years to perform tissue elasticity imaging are reviewed extensively by several authors [18]–[21]. Here, before briefly exploring specific methods, the theoretical basis of elasticity imaging is formulated.

1.2.1 Soft tissue biomechanics

In this section, several concepts from soft tissue biomechanics are provided to describe the general approach taken when obtaining tissue visco-elastic properties in ultrasound elastography.

Stiffness is the extent to which an object resists deformation and exhibits as a response to the stress applied to the tissue. In general, soft tissues can be described as a visco-elastic medium, thus exhibiting properties of both elastic solids and viscous fluids, while particular behavior depends on the frequency of excitation. Its response under stress can be approximated using the models shown in Fig. 1.2.

In a purely elastic medium (Fig. 1.2(a)) the stress σ (equal to external force per unit area) and strain ε (equal to expansion per unit length) are proportional. This
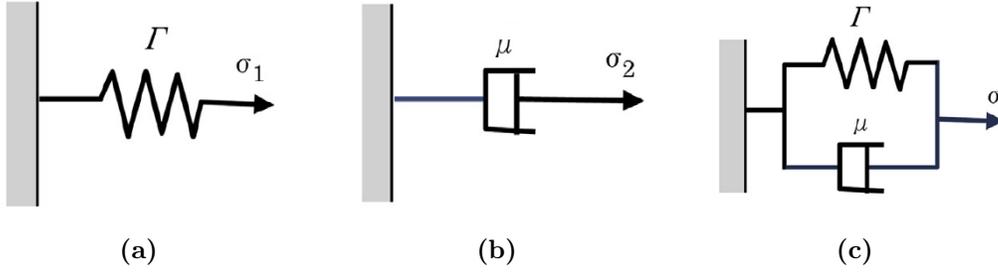
Figure 1.2: Visco-elastic model of the tissue: (a): Elastic component. (b): Viscous component. (c): Kelvin-Voigt model of the tissue that combines elasticity with viscosity.

relationship is known as Hooke's law and proportion is described by elastic modulus Γ:

$$\sigma_1 = \Gamma \cdot \varepsilon \tag{1.1}$$

For a viscosity component (Fig. 1.2(b)), the stress increases along with the speed of deformation — strain rate $d\varepsilon/dt$. The coefficient that is a parameter of this proportionality is the viscosity coefficient μ:

$$\sigma_2 = \mu \cdot \frac{d\varepsilon}{dt} \tag{1.2}$$

In practice, soft tissues are complex structures which mechanical characteristics are determined by a complex combination of both elasticity and viscosity. Various models can be used to approximate real tissue behavior [22]. Fig. 1.2(c) presents the commonly used Kelvin-Voigt model, in which strain and stress are related by a combination of Eq. (1.1) and Eq. (1.2):

$$\sigma = \Gamma \cdot \varepsilon + \mu \cdot \frac{d\varepsilon}{dt} \tag{1.3}$$
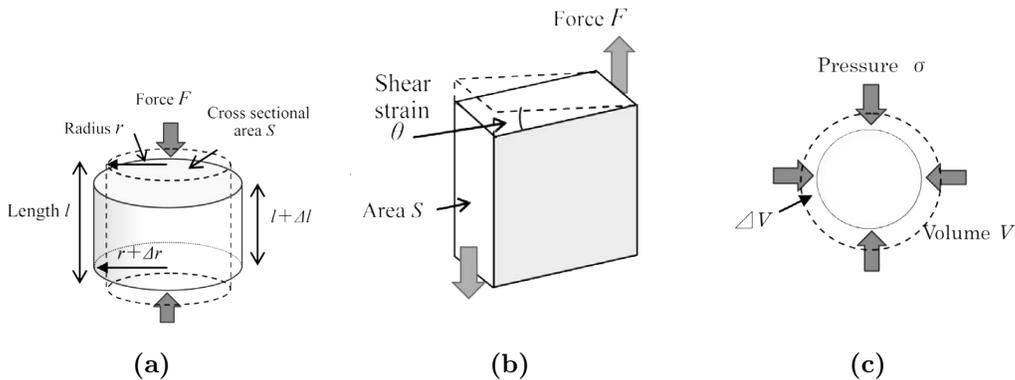
Figure 1.3: Various elastic moduli: (a) Youngs's modulus, (b) Shear modulus, (c) Bulk modulus. Source: [23].

For elastic component, depending on the deformation type, there can be three different elastic moduli defined to describe the resistance of material to deformation. Those are introduced below.

19

- In the literature, the stiffness of a material is most commonly reported in terms of Young's modulus E. It defines the ratio between axial stress $\sigma = F/S$ (see Fig. 1.3(a)) — axial force per cross-sectional area) and longitudinal strain $\varepsilon_L = \Delta l/l$:

$$\sigma = E \cdot \varepsilon_L \tag{1.4}$$

In a case where there is no change in volume and in response to axial stress both longitudinal strain and transverse strain $\varepsilon_r = \Delta r/r$ can be observed. The ratio between those strains is called Poisson's ratio:

$$\nu = \frac{\varepsilon_r}{\varepsilon_L} \tag{1.5}$$

- Shear modulus G is related to deformation in which parallel internal surfaces slide past one another. It defines a ratio between shear stress $\sigma = F/S$ (see Fig. 1.3(b)) and shear strain $\varepsilon_S = \theta$:

$$\sigma = G \cdot \varepsilon_S \tag{1.6}$$

- Bulk modulus K is the inverse of compressibility — it quantifies the resistance to volume strain $\varepsilon_V = \Delta V/V$ (see Fig. 1.3(c)) in response to increase of pressure:

$$\sigma = K \cdot \varepsilon_V \tag{1.7}$$

Assuming a homogeneous, isotropic, linear elastic material all these moduli define the elasticity of the medium and are related as below [24]:

$$E = \frac{G(3K + 2G)}{K + G} \tag{1.8}$$

$$\nu = \frac{K}{2(K + G)} \tag{1.9}$$

$$G = \frac{E}{2(1 + \nu)} \tag{1.10}$$

Since the water content in soft tissues is high, those can be assumed as incompressible — $\nu \approx 0.5$, and consequently, formula (1.10) simplifies to:

$$E \approx 3G \tag{1.11}$$

All the coefficients introduced above besides being essential for the description of the static deformations, it turns out, that they also determine the conditions for mechanical waves propagation — their propagation speed in specific. There are two primary modes of wave propagation in soft tissues. In longitudinal waves, particles oscillate in the direction that the wave propagates. Adversely, in shear waves, particle oscillations occur transversely to the wave propagation direction. Equations modeling wave propagation can be derived from the constitutive properties of the medium and Newton's 2nd law. The general linear equations of dynamic equilibrium describing the motion of a mechanical body in Cartesian coordinates can be expressed in terms of Cauchy's stress tensor $\boldsymbol{\sigma}$:

$$f_i = \frac{\partial \sigma_{ij}}{\partial x_j} = \rho \cdot \frac{\partial^2 u_i}{\partial t^2}, \tag{1.12}$$

where $\vec{f}$ is externally applied force per unit volume, u_i are the components of the displacement vector and t is time. The formula for infinitesimal strain tensor components can be written as [18]:

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{1.13}$$

For complex materials like tissues, it is nearly impossible to derive a constitutive equation that describes a relation between stress and strain accurately at any loading [24]. Therefore, typically certain assumptions are made and tissues are often described to a first-order approximation as linear and elastic solids, ignoring this way viscous forces and non-linearity [20], [24]. In addition, tissues are also assumed to be isotropic and homogeneous and then the relationship between stress and strain can be expressed as follows [20]:

$$\sigma_{ij} = \lambda\varepsilon_{kk}\delta_{ij} + 2\mu\varepsilon_{ij}, \tag{1.14}$$

where λ and μ are Lame coefficients and δ_{ij} is the Kronecker symbol. Equations (1.12)–(1.14) can be combined to obtain a formula in terms of the displacement vector alone [20]:

$$(\lambda + \mu)\frac{\partial^2 u_j}{\partial x_j \partial x_i} + \mu\frac{\partial^2 u_i}{\partial x_j \partial x_j} = \rho\frac{\partial^2 u_i}{\partial t^2}. \tag{1.15}$$

By using the Helmholtz theorem, one can decompose the displacement vector field into components related to longitudinal and shear wave propagation [24]. In specific, shear waves propagation can be expressed as [20]:

$$\nabla^2\vec{u} = \frac{1}{c_s^2} \cdot \frac{\partial^2\vec{u}}{\partial t^2}, \tag{1.16}$$

where $\vec{u}$ is the displacement vector, ∇ is the Laplacian operator determining the spatial gradients of the displacement vector, t is time and c_s is the shear wave propagation speed that typically ranges between 0.5 to 10 m/s in soft tissues and equals:

$$c_s = \sqrt{\frac{G}{\rho}}. \tag{1.17}$$

Longitudinal wave propagation speed c_l can be found in a similar way — details can be found elsewhere [25]. Assuming $K >> G$, it can be expressed as:

$$c_l = \sqrt{\frac{K + 2G}{\rho}} \approx \sqrt{\frac{K}{\rho}}. \tag{1.18}$$

For soft tissues, c_l is frequently assumed to be approximately 1540 m/s with small differences for different types of tissue structures.

Combining formulas (1.11) and (1.17) the following simplified formula can be obtained, which shows that tissue elasticity properties can be estimated using shear wave propagation speed. This builds the main concept behind shear wave elastography:

$$E \approx 3\rho \cdot c_s^2. \tag{1.19}$$

In contrast to longitudinal waves, shear waves of high frequency (MHz range) are attenuated strongly in the tissues, and the energy of shear waves propagating through

the medium occupies the band below 1 kHz typically, what results in low propagation speeds. However, when the frequency of shear waves increases, the effect of velocity dispersion due to viscosity component starts to be significant. For example, for the Kelvin-Voigt model, rather than using formula (1.17), an alternative formula for shear wave speed can be used to take into account the viscosity of the soft tissue [26], [27]:

$$c_s = \sqrt{\frac{2\left(G^2 + (2\pi\mu f)^2\right)}{\rho\left(G + \sqrt{G^2 + (2\pi\mu f)^2}\right)}}, \tag{1.20}$$

where the μ is the shear viscosity of the medium and f is the shear wave frequency.

The difference between c_l and c_s in values is in order of several magnitudes. Thus, it creates a technical opportunity to track the propagation of shear waves through the medium using pulse-echo ultrasound. In addition, big differences in shear wave propagation speeds for various tissues allow tissue elasticity imaging with high contrast.

The analysis presented above made certain assumptions by treating tissues as a medium that is linearly elastic, homogeneous, isotropic, infinite, and continuous. Tissues are complex structures and those assumptions are not entirely true in practice:

- The relation between force and deformation is usually nonlinear [28] (see Fig. 1.4). Tissue stiffness increases with stress increase. For strain elastography, the greater the pre-stress is, the smaller will be the achieved strain for a given additional applied force. For shear-wave methods, shear wave speed increases along with the pre-stress.

- Tissues exhibit viscosity along with elasticity. Thus, the stress-strain relation is time-dependent. Due to viscosity, materials stiffen the faster they are strained. In the case of SWE, the use of higher shear wave source frequencies will result in higher values of estimated shear wave speed [26].

- Elastic properties of some tissues vary spatially and directionally. Some examples of tissues that were reported to exhibit anisotropy in SWE are muscle [29], skin [30] or kidney [31], [32].

- Tissue boundaries and structure discontinuity may change the relationship between shear wave speed and shear modulus, since the shear wave may propagate as a guided wave that follows the tissue structure. The most distinct example is shear wave propagation through arterial walls [33]. A proper correction method shall be applied in order to eliminate possible artifacts due to this phenomenon.

The limitations above should be taken into consideration upon an interpretation of obtained results since all these can produce artifacts or bias. Ongoing research aims to take account of them, gain additional information by incorporating the above limitations into new methods, and afford elastography the ability to image new tissue characteristics.

1.2.2 Acoustic radiation force

Acoustic radiation force (ARF) is a phenomenon related to the propagation of acoustic waves through a dissipative medium. As a result of absorption and reflection of the wave, there is a transfer of momentum from the wave to the medium in which it propagates [34].
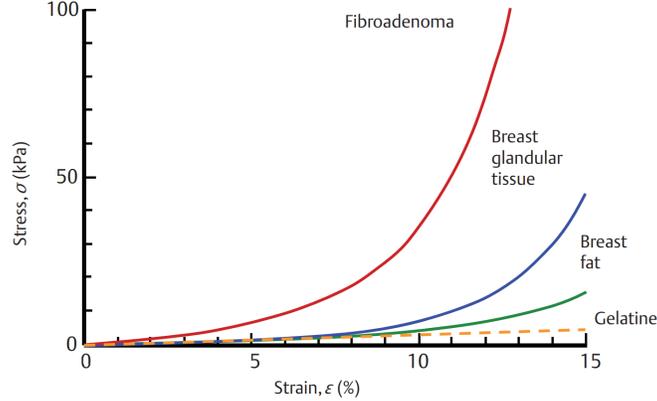
Figure 1.4: The relationship between stress and strain plotted for a linear elastic medium (gelatine, 7.3% by weight) and several different tissues exposing nonlinear behavior. Young's modulus is the slope of the curves shown. Source: [23].

This transfer results in the body force application which acts in the direction of wave propagation. For an exponentially attenuating plane wave, this force can be expressed in terms of the time-averaged intensity $\vec{I}$ at any given spatial location as [35], [36]:

$$\vec{F} = \frac{2\alpha\vec{I}}{c_L} \tag{1.21}$$

where F is acoustic radiation force, c_L is the longitudinal wave propagation speed in the medium, α is the absorption coefficient of the medium, and I is the temporal average intensity of the beam at a given position. The idea of employing ARF for the assessment of tissue mechanical properties goes back to Sugimoto et. al. [37], where impulsive ARF was used to generate localized force within the tissue and resulting deformation was measured using pulse-echo ultrasound. This concept was later utilized by the group at Duke University led by Nightingale in a series of works [35], [36], [38] and was named acoustic remote palpation. The feasibility of using ARF to induce tissue displacements remotely for imaging the mechanical properties of tissues has been demonstrated using a standard diagnostic transducer with phantoms [35] and *in vivo* [36].

In order to induce detectable tissue displacements, ARF strong enough needs to be generated. Therefore, elastography methods employing ARF use focused beams and long, high-power excitations. A temporally impulse-like (i.e., short duration, from tens of us up to 1 ms) ARF generated with a focused acoustic beam is called acoustic radiation force impulse (ARFI) or push pulse. The displacement field resulting from the application of ARFI is spatially distributed and dependent upon the local medium properties and characteristics of the transmitted beam. Therefore, displacement magnitude distribution is unknown in advance in *in vivo* applications. As shown in Eq. 1.21, the ARF is a function of the absorption and intensity. Although higher attenuation increases the amount of momentum transfer to the medium, at the same time it decreases the intensity of the acoustic wave on the path to a given depth. Because attenuation is frequency-dependent, the frequency used to generate an ARF should be optimized for a specific application. Peak ARF magnitudes are typically on the order of dynes in *in vivo* applications, creating tissue displacements in the range of 1 to 10 um [24]. Softer tissues move farther, take longer to reach a peak displacement, and recover more slowly than
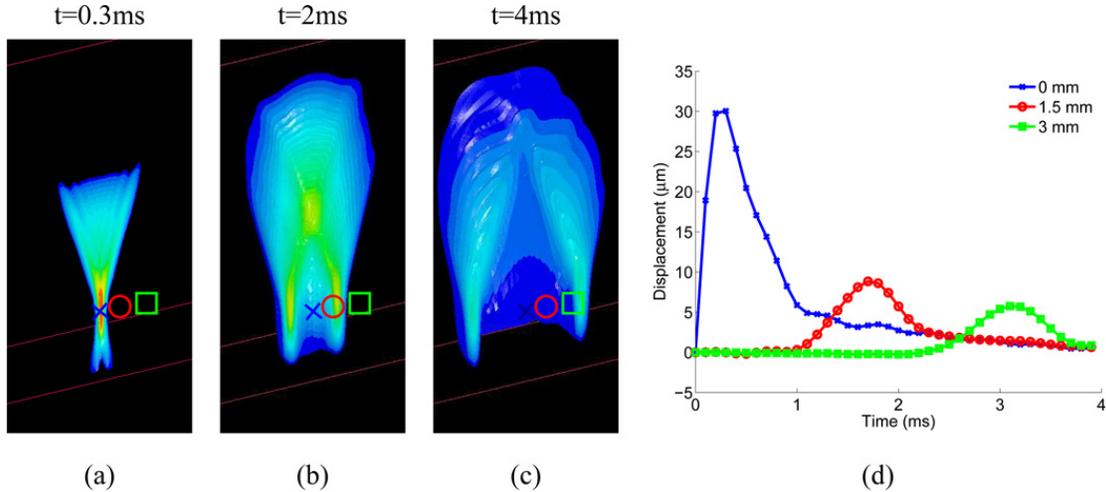
Figure 1.5: (a)–(c): Simulated axial displacement response in a 3-D, homogeneous, isotropic, elastic tissue (Young's modulus = 4 kPa) to an ARFI configured as focused at 20 mm (F/1.3 focal configuration, frequency = 6.7 MHz), impulsive and 50 us long. A transducer is placed at the top of the images, focused in the center of the plane (marked by the blue X). The images show the central axial/lateral plane that is aligned with the transducer. (d): The displacement at each symbol (x, circle, square) as a function of time after excitation. Positive displacement means displacement in the direction of acoustic wave propagation. Initially, in (a), the axial displacement response reflects the spatial distribution of the applied ARF field. After that, shear waves propagate out of ROE and the amplitude of oscillations drops with propagation distance due to geometric spreading and attenuation. Image source: [39].

stiffer tissues [36], [39]. It should be noted, that in conventional ultrasound imaging, the ARF phenomenon also occurs, however, due to a short length of transmit pulses resulting displacements are negligible.

An important effect associated with an application of localized and impulsive ARF is a generation of shear stresses within the tissue which induces a transient shear wave that propagates away from the region of excitation (ROE) in the directions perpendicular to the ARF [40], as shown in the Fig. 1.5. Since the shear wave propagation speed is related to mechanical properties of the medium (see Eq. 1.19, this effect plays an important role in methods that rely on the estimation of tissue stiffness using shear wave propagation speed measurement. The frequency content of generated shear waves is modulated by the size of the focal area of a pushing beam, the length of excitation, and the stiffness of the medium [41].

1.2.3 Methods

Since the beginnings of elastography in the early 1990s, a wide variety of ultrasonic methods have been investigated to obtain information about the mechanical properties of the tissues. The intention of this section is to briefly describe the most relevant methods to present the position of 2-D shear wave elastography which is the topic of this dissertation. For a more comprehensive review of elastography methods please refer to [18], [28], [42].

All the elastography methods determine tissue elasticity properties imaging indirectly — each involves two common elements: application of force and measurement of the local mechanical response of the tissue under stress. By way of excitation force

application, elastography methods can be classified as those using mechanical excitation (manual compression, actuators) and those using ARF. In addition, excitation can be applied quasi-statically, transiently (impulsive), or harmonically. In the measurement part, methods can be further classified by the property displayed for those displaying strain, strain rate, displacement, or shear wave speed (SWS). Finally, each method can provide qualitative or quantitative information and can perform either point measurement or reconstruct the elasticity image within a ROI or in the whole FOV of the probe. Table 1.1 shows how the most relevant methods can be classified using those criteria.

Table 1.1: Summary of ultrasound-based elastography methods.

Method	Excitation type	Measurement locations and parameters
Vibration Sonoelastography	Mechanical actuator; harmonic	2D displacement distribution obtained in whole FOV
Strain Elastography	Manual compression; quasi-static	2D strain image obtained in whole FOV or in ROI
ARFI Imaging	ARF; transient	On-axis displacement
Transient Elastography	Mechanical actuator; transient	Shear wave speed
Shear Wave Elasticity Imaging (SWEI)	ARF; transient	Off-axis shear wave speed
Supersonic Shear Imaging (SSI)	ARF; transient	Off-axis shear wave speed
Spatially Modulated Ultrasound Radiation Force (SMURF)	ARF; transient	Off-axis shear wave speed
Vibro-acustography	ARF; harmonic	On-axis displacement
Harmonic Motion Imaging	ARF; harmonic	On-axis displacement
Shearwave Dispersion Ultrasound Vibrometry (SDUV)	ARF; harmonic	Off-axis shear wave speed
Crawling Wave Spectroscopy (CWS)	ARF; harmonic	Off-axis shear wave speed

One of the first successful forms of elasticity imaging was **vibration sonoelastography**. In the first studies held by Lerner and Parker [43] and Krouskop et al. [44] a low-frequency vibration (20—1000 Hz) was externally applied by a motorized actuator or an acoustic horn to excite internal vibrations within tissue under inspection. Depending on the shape of the tissue or organ, its structure, and elastic boundary conditions, tissue resonated and formed multiple reflected modes. An induced motion was measured using Doppler techniques with an ultrasound transducer placed in proximity to the vibration source. The amplitude of the vibration was smaller in the hard tissue than in the soft tissue which gave means for tissue elasticity differentiation. This method was further analyzed and developed by the group of Lerner and Parker resulting in real-time tissue elastic constant measurements, sonoelastography finite element modeling results [45], phantom, *ex vivo* [46] and *in vitro* [47] tissue results. Yamakoshi et al. developed a vibration phase gradient approach to sonoelastography [48], which allowed them to derive the wave propagation velocity along with dispersion properties for the assessment of both elastic and viscous characteristics of the tissue. A more recent approach to vibration sonoelastography involved using two mechanical actuators instead of one to generate shear wave interference patterns called "*crawling waves*" [49], [50], which motion could be tracked with an ultrasound scanner's standard frame-rate (less than 50 frames/s).

The most widespread form of elastography so far, and also the first one put into

practice, was **compressional** or **strain elastography**, initiated by Ophir et al. in his seminal work [51]. This conceptually simple approach is also named quasi-static strain imaging. It relies on extracting tissue elasticity information by acquiring maps of anatomy before and after a small compression which is done manually by pushing the probe towards a body. The whole process principle is diagrammed in Fig. 1.6. Typically, the before and after compression radiofrequency (RF) echo signals are compared by a speckle-tracking algorithm in order to measure the amount of local deformations, i.e., the strain of a medium. The most common methods of motion tracking in RF echo signals are correlation-based algorithms. Softer tissues tend to deform more than harder tissues. Therefore, by differentiating the displacements at spatial positions, local strains can be found that then build the strain image. Many research groups through the years have contributed to this technique by bringing signal processing improvements. Review articles that detail developments in this field and procedural guidelines are available [52], [53].
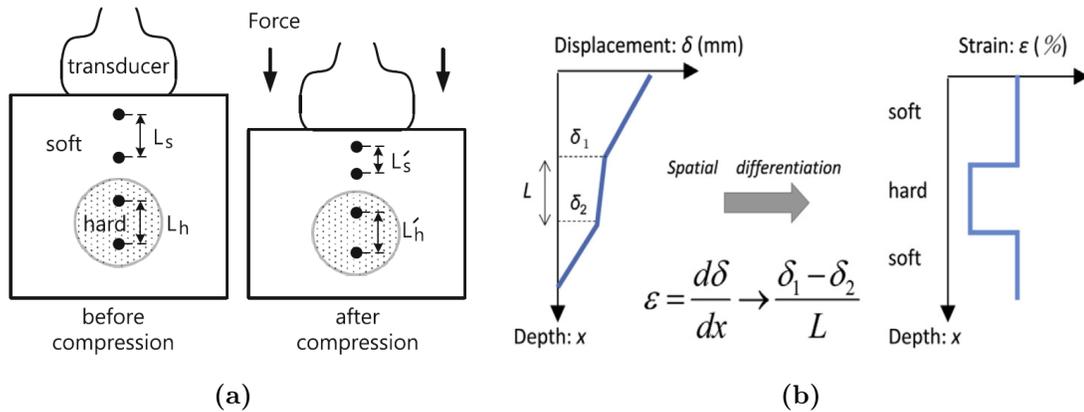
Figure 1.6: Principles of strain elastography: (a) Echo data are acquired from a medium of interest before and after applying slight compression (typically around 1%). The scatterers in the medium undergo different displacements depending on local stiffness. In this case strain within soft region is greater than in stiffer region: $(L_s - L'_s)/L_s > (L_h - L'_h)/L_h$. Those displacements can be obtained at each site by calculating the correlation between the echo signal before and after compression. Source: [21]. (b) Strain ε is obtained by spatial differentiation (gradient) of displacement — that is — a ratio of the difference in displacement between any two points to their distance before compression, L. Source: [23].

Theoretically, a local map of Young's modulus E can be obtained if local stress and strain are known, however, because it is difficult to actually calculate the stress distribution *in vivo* [23], it is assumed to be uniform and only a relative stiffness map can be estimated. The advantage of strain elastography is that the handheld ultrasound scanning transducer can be used to produce a localized compression near the ROI in the breast and other applications, particularly in more superficial targets of interest. The first commercial ultrasound system implementing strain elastography became available in 2003. For nearly 20 years many manufacturers have produced ultrasound systems supporting strain elastography and the clinical usability of this well-established technique has been shown in many applications [53].

Another method — **ARFI imaging** — capable of imaging local variations of mechanical properties of tissue was developed by the group of Nightingale et al. [35], [36]. This method utilized ARF to apply localized displacements deep within the tissue. The

same transducer was used to both induce deformation and detect transient tissue response in an acquisition that consists of three types of pulses at each lateral position. Reference pulses are used to capture the reference position of the tissue before the application of ARF. Then, a push pulse is generated to induce tissue displacement, and immediately after that, in a third step, a tracking pulses sequence is generated to measure an on-axis response and recovery of the tissue within the ROE. 2-D image data are generated by sequentially interrogating different lateral positions along the array aperture (Fig. 1.7(a)). Axial displacements through time data can be obtained from tracking pulses using correlation-based or phase-shift algorithms. This data can be used to build an image that depicts the relative stiffness within ROI. Various data types can be used to create an image: displacement at the specified time after pushing excitation, maximum tissue displacement, or time to maximum displacement [36]. In general, softer regions will displace farther, take longer to reach a maximum displacement and recover more slowly than stiffer tissues [24].

The ARFI imaging technology has been implemented commercially by Siemens (Munich, Germany) [54] and has been investigated in many clinical applications using various transducers and beam sequences. A comprehensive review that covers various aspects of this technique was written [39].

Unlike strain elastography and ARFI imaging, which rely on measuring tissue response to induced displacement, there was a group of methods developed based on monitoring the shear wave propagation in tissue. These methods estimate the tissue stiffness quantitatively by measuring shear wave propagation speed, which is, under certain assumptions - related to underlying tissue stiffness by Eq. (1.19). A number of methods have been developed exploiting various shear waves generation sources, tracking methods, and data processing approaches.

The first technique using this approach was **transient elastography**, which used an external actuator to apply a mechanical shock to soft tissue within a short time interval and ultrasound methods to track the resulting motion. This type of excitation generates four types of waves including a compressional and shear wave in the medium. Due to big differences in propagation speeds, the compressional and shear waves could be separated, what was shown in a study held by Catheline et al. [55]. The tissue motion caused by excitation was observed using high frame-rate imaging. Firstly, the one-dimensional (1-D) measurement technique was developed [56]. Unlike the conventional method, in which the ultrasonic transducer and the low-frequency actuator were two separate parts, that method relied on a probe that integrated the vibrator with the transducer, which was built on the axis of the vibrator. Nearly at the same time, the two-dimensional (2-D) imaging technique was also developed by the same group [57]. For this purpose, a dedicated system was developed which contained a vibrating device surrounded by a linear array of 128 transducers that performed high frame-rate imaging (up to 10,000 frames/s) to track the propagation of shear waves in real-time. Since 2001, the company Echosens (Paris, France) has been commercializing a 1-D transient elastography device under the name *FibroScan*, which is used for the assessment of liver hepatic fibrosis [58].

In **shear wave elastography (SWE) methods**, instead of a mechanical actuator, the ARFI is used to create shear waves within the medium. The same transducer that generates ARF is used to apply the ARF and to track generated shear waves outside the ROE using conventional pulse-echo ultrasound imaging. This approach was initially described theoretically by Sarvazyan et al. [40] and investigated experimentally by the group of Nightingale [59]. These works were followed by many investigations

27

and developments, that differed in pushing beam design, ARFI duration (transient or harmonic), and shear waves tracking beam design and positioning.
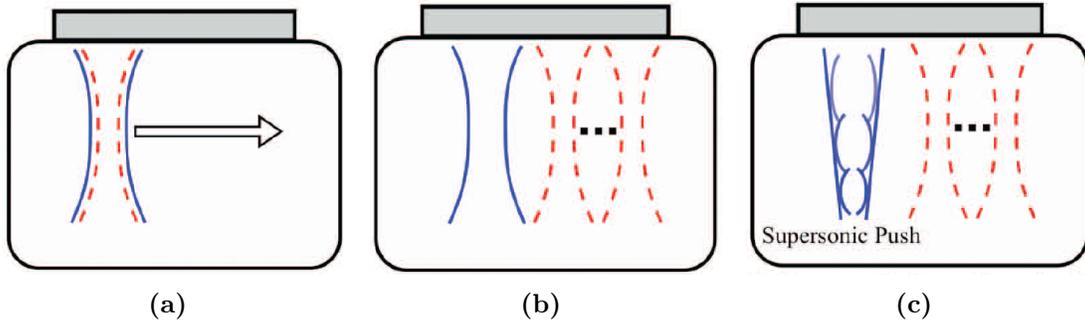
Figure 1.7: Principles of selected elasticity imaging methods based on ARF generation. The same transducer (shown on top of each diagram) is used to generate push pulse (blue beams) and to detect resulting displacements (tracking beam locations shown in red dotted lines). (a): ARFI imaging — tracking beam is aligned with the pushing beam. The push and tracking sequence is repeated and swept across the lateral dimension. (b): SWEI — shear waves are generated by a single focused push pulse and then tracked on multiple off-axis lateral positions away from ROE. (c): Multiple successive push pulses focused at multiple axial depths are generated to create shear wavefronts. Shear wave propagation is monitored in multiple locations using plane wave imaging. Images source: [24].

Nightingale et al. in [59] used a single focused beam to generate ARF, whereas shear waves propagation tracking with sufficient frame rate was accomplished by repeating the push pulse and sequencing tracking beam locations across the ROI. This method, called **shear wave elasticity imaging (SWEI)** allowed to obtain displacement vs. time data at multiple lateral locations (see Fig. 1.7(b)), such that shear wave propagation speed could be estimated. In this early work, a direct inversion of Eq. (1.16) was employed for this purpose. Later on, due to large jitter, the method was modified to determine local shear wave speed using time-of-flight (ToF) methods and successfully commercialized by Siemens (Munchen, Germany) in 2008 [54], and has been used in many clinical applications like the noninvasive assessment of liver fibrosis [60]. However, as a result of focusing, the shear waves were generated in close proximity to the focal point of the pushing beam. Therefore, high-quality elasticity map reconstruction was possible only in limited depth-of-field (DOF) or in a small ROI that allowed only a point-like measurement. To address this issue, a new push beam generation method was proposed by Bercoff et al. in [61] — **Supersonic Shear Imaging (SSI)**. In SSI, a sequence of beams focused at consecutive depths was used to generate broad shear wavefronts over the high axial extent to cover a big part of FOV. In addition, in this method, high frame rate imaging of several kilohertz frames-per-second (FPS) was utilized for shear wave tracking. This high frame rate was achieved by eliminating focusing when transmitting pulses used for motion detection. Instead, one plane wave or a set of plane waves with different angular directions were transmitted for tracking the shear wave propagation. It allowed shear wave real-time propagation capture in the whole FOV within a single acquisition scheme, which was earlier introduced by the same group in [62]. Although a single SSI acquisition covered a big part of FOV, there was still a lack of shear waves in the beam axis. Therefore, the authors used 3 acquisitions, each with different pushing beams' lateral position [63]. The SSI was implemented and brought to the market by Su-

perSonic Imagine (Aix-en-Provence, France) and used in many applications, like breast cancer lesions detection [63], mapping hepatic viscoelasticity [64], stiffness assessment in musculoskeletal [29] and cardiovascular [65] systems.

For nearly two decades SWE methods have been in the focus of the ultrasound research community and have been subject to improvements that covered various aspects of this technique. Researchers tried to optimize push pulse generation and data processing algorithms to improve the overall quality of obtained images. Some efforts in this field are described in chapter 2 of this thesis. Moreover, some sophisticated techniques have been implemented that benefit from using ARF to induce shear waves in the tissue. In **Spatially Modulated Ultrasound Radiation Force (SMURF)** [66] instead of tracking shear waves propagation at multiple locations generated by a single push beam, two separated push excitation locations are used and shear wave is detected at a single tracking location to measure the difference of time-of-arrival of the resulting shear waves. SMURF is less susceptible to the varying biases that exist at multiple tracking locations ToF estimations used in other methods [67]. This advantage can lead to contrast and resolution gains in some applications [68].

A number of methods were developed that instead of impulsive ARF excitation, apply push pulses harmonically to excite the tissue at specific frequencies. **Vibro-acustography**, a method developed by Greenleaf and Fatemi [69] generated vibration of an object at the intersection of two focused pushing beams of slightly different frequencies. Such excitation resulted in the low-frequency acoustic response (tens of kHz) of the object that was registered by a hydrophone. This response was dependent upon mechanical properties of the tissue. An image could be formed by translating the position of the pushing beams' focal point across the ROI. A more recent approach to this technique was presented by Urban et al. [70] who adopted this method for a clinical ultrasound system. A review of various aspects of this technique is provided in [71]. In a modification of this technique, **Harmonic Motion Imaging (HMI)**, Konofagou et al. proposed to utilize the same concept of overlapping pushing beams, but with the use of a separate transducer to monitor the induced tissue motion [72]. This modification was found especially convenient for HIFU (high intensity focused ultrasound) therapy thermal ablation monitoring [73], [74]. A review of this technique along with its applications is covered in [75].

Shearwave dispersion ultrasound vibrometry (SDUV) [27] used a repetitive push pulse at the same spatial location to produce harmonic shear waves with designed frequencies to assess the dispersion of viscoelastic medium. The phase difference of the shear wave between two lateral locations along its propagation path was used to calculate shear wave speed within the tissue. By applying this over a range of excitation frequencies, a quantitative estimate of tissue stiffness and viscosity could be derived using a model-based approach. This solution was implemented and commercialized by Philips (Amsterdam, The Netherlands). In another method developed by the group at the University of Rochester, **crawling wave spectroscopy (CWS)**, authors used two pushing beams of slightly different frequencies and focused at two points separated in lateral dimension to create two shear sources [76]. Each push beam was generated harmonically at slightly different repetition frequencies. It implemented the concept derived previously in [49] for sonoelastography but using harmonic ARF rather than mechanical excitation. As a result of such interfering shear wave sources, the slow-moving "crawling wave" was generated that could be tracked in real-time using Doppler techniques. The speed of the crawling wave was related to the underlying mechanical

properties of the tissue giving means for qualitative stiffness estimation. The same group has successfully integrated CWS on a commercial ultrasound system [77], [78].

As shown above, although aiming similar objective, each of the ultrasound-based elastography methods provides a unique mechanism to obtain information about the mechanical properties of the tissue. Consequently, each method has its own advantages and disadvantages. Out of all the methods covered, vibration sonoelastography and transient elastography use mechanical actuators. Although it may lead to large motion amplitudes that can be detected and measured more easily, a requirement of additional hardware makes them inconvenient. Admittedly, tightly coupling the transducer with the actuator solves this problem in the case of transient elastography, but this method has very limited applications. Sonoelastography is considered disadvantageous because of the need for indirect observation of the vibration pattern, what makes it difficult to obtain a high-resolution distribution of stiffness [21]. The strain elastography technique, due to its simplicity (both conceptually and in terms of implementation effort) is widely spread in radiology. However, the main limitations of this technique are still the control of the stress applied, which is not uniform and remains operator dependent, and the absence of a quantification [42]. In addition, the use of stress applied by the operator limits the technique to superficial organs, mainly the breast or the thyroid. In the rest of the methods described the ARF is used as the excitation. It is advantageous because the same transducer can be used to induce localized motion deep into tissue and to track it. In addition, the excitation applied is not operator-dependent. Nevertheless, allowable ultrasound output intensity is limited to avoid mechanical and thermal bioeffects [79] which results in ARF-induced motion amplitudes in the range of microns. Detecting such small displacements accurately is challenging and limits the image quality. In particular, it is difficult to induce detectable displacements at depths below $6 \, \text{cm}$ [18], what limits the depth of ARF-based imaging methods. Still, a generation of ARFI requires robust system capabilities, that involve transmit capabilities tailored for ARF generation [80]. The main advantage of methods based on the estimation of shear wave propagation speed is that they provide both qualitative and quantitative results. It should be also noted, that some SWE methods (SDUV, CWS) are designed specifically to obtain viscoelastic properties of the tissue rather than only elastic characteristics, what is achieved by temporal modulation of ARF excitation to control the frequency of applied stress. Alternatively, shear wave dispersion needs to be explored from an acquisition that involves impulsive ARF excitation as in SWEI or SSI, some dedicated data processing in Fourier domain can be exploited to extract signal components at specific frequencies [26], [81].

1.2.4 Medical applications

Most of the aforementioned methods have been used for *in vitro*, *ex vivo*, or *in vivo* measurement of soft tissues mechanical properties. While some of them have been successfully implemented in commercial systems, numerous reports were published exploring various techniques in various applications, such as: assessment of liver fibrosis [82], [83], breast cancer screening [84], prostate abnormalities screening [85], thyroid nodules evaluation [86], lung disease assessment [87] and many others. Elastography has been also used in cardiovascular system diagnostics by measuring arterial wall stiffness [33], [65] or for monitoring the ablation in HIFU treatment [73], [74], [88]. Based on this experience, general guidelines for ultrasound elastography methods and more specific

guidelines for selected organs imaging have been developed [23], [89]–[93].

It should be noted, that different methods may have particular advantages in different tissues or applications. For example, one should distinguish the task of characterizing stiffness of a region of tissue when there is a lack of background to provide contrast (i.e., liver fibrosis staging), from that to generate images of elasticity of stiffer inclusions (i.e., breast lesion visualization). In the former task, it is beneficial to use SWE techniques, since it is required to obtain a quantitative result. Generally, qualitative techniques such as strain imaging or ARFI imaging, HMI or vibroacustography have improved spatial resolution in comparison to quantitative methods including SWEI, SSI, SMURF, SDUV, and CWS [24]. Therefore in the latter case, for the task of imaging inclusions, qualitative methods may be better suited. However, quantitative methods are still able to perform imaging of inclusions, even providing quantitative estimates of stiffness but possibly at the cost of reduced resolution.

1.3 Thesis motivation

Portable point-of-care ultrasound has been emerging as a new ultrasound paradigm due to its unique advantages including small size, lightweight, fast startup, and affordability. Modern portable battery-powered ultrasound devices can offer high-quality imaging and even advanced ultrasound modes like color Doppler and allow a real-time examination wherever the patient is present. The use of these portable and pocket-sized ultrasound devices has been studied and proven practical in many medical specialties and use cases [14]. Low unit cost combined with reasonable imaging quality of portable devices in comparison to cart-based systems is considered a key factor that may extend the use of ultrasound imaging in lower resource environments or in primary care units, where availability of ultrasound diagnostics is limited [94]. In specific, in austere environments, ultrasound may be the only available imaging modality and thus can guide diagnosis or provide screening. However, by the time of writing this text (2023), elastography modes in portable ultrasound devices with one exception are not available. Therefore, sites in which ultrasound is brought by point-of-care devices can not benefit from applications in which elastography is well known for its usefulness for diagnostic and screening purposes.

In the point-of-care class of devices, there is a set of well-desired features including: low power consumption for a long time of operation on a built-in battery and small heat dissipation, portability, and low unit cost. Therefore, there are a lot of efforts put into power optimization of both system hardware and imaging scenarios. In this context, ultrasound-based elastography methods detailed in the previous section vary in terms of transferability to the portable class of devices. Methods based on mechanical excitation are inconvenient as an external actuator is required, which also needs a dedicated driving circuit. Applications of these techniques are also quite limited. On the other hand, strain elastography has unique features that make this technique implementable using portable devices. It does not involve any electric power for excitation generation, since the stress is applied by the operator. In addition, it also does not require high framerate acquisition. Although the elasticity map reconstruction algorithm is complex and computationally intensive, there were researches aimed to propose efficient algorithms suitable for portable ultrasound [95]. In fact, strain elastography technique was successfully implemented on hand-held pocket size Clarius L7, C3, and EC7 scanners (Clarius,

Vancouver, Canada [96], [97] and they still remain the only point-of-care devices that feature elastography mode.

While manual compression as an excitation is beneficial in terms of transferability to a portable device, at the same time it is operator-dependent and allows only qualitative imaging of superficial organs. Since the lack of proper training is considered to be one of the main factors limiting the spread of point-of-care ultrasound [15], quantitative results less dependent on the operator brought by shear wave elastography may make image interpretation easier. Moreover, the well-round SWE technique offers a broader applications portfolio than strain elastography.

However, practical implementation of the SWE technique is a complex engineering task and poses a number of technical challenges. Firstly, a generation of ARFI involves the transmission of high-energy, long transmit sequences using a high number of probe elements. Consequently, it requires robust transmit capabilities of the ultrasound system and can cause significant heat dissipation by the probe driving electronics and the probe itself. For this reason, SWE methods utilizing repetitive harmonic ARF excitation seem impractical for portable devices. Secondly, the shear wave propagation tracking requires high frame-rate imaging — minimum 1000 fps, and typically more than 5000 fps. This leads to a significant amount of data being captured and transferred out for further processing. A huge dataset to be processed in the software and the complex nature of the elastogram reconstruction algorithms is computationally intensive and requires additional hardware resources such as dedicated CPUs or GPUs in order to provide sufficient performance to reconstruct elasticity maps, especially in real-time applications. Due to the aforementioned reasons, SWE is less suitable for portable implementation and so far has been reserved for high-performance premium ultrasound systems.

The 2-D SWE was successfully developed by several industry companies and was made commercially available. Table 1.2 lists those companies with related device models. All these devices are expensive premium cart-based systems. Nevertheless, most mentioned devices are clinical systems and although are useful for medical diagnostics, their imaging implementation is hidden from the user, and algorithm control or customization is not provided in any way. For this reason, their usefulness in medical ultrasound research is limited. To effectively facilitate the development and evaluation of new diagnostic ultrasound methods and algorithms, researchers require a dedicated system, that features full configurability of transmit patterns and acquisition sequences, provides access to raw echo pre-beamformed receive data (RF data), and has sufficient computational resources to allow real-time imaging implementation [9]. In addition, to support SWE technique, such a research platform should also allow generation of long transmit bursts for push beam generation. To date, extended options of Vantage by Verasonics (Kirkland, WA, USA) and Prospect T1 by S-Sharp (New Taipei City, China) are the only available research scanners that feature all these functionalities. Nevertheless, these scanners are bulky and expensive.

During the last decade, extensive development in the field of ultrasound systems design has been realized in the group of Lewandowski et al. ([80], [98]–[102]) at the Institute of Fundamental Technological Research of the Polish Academy of Sciences. Some of those efforts resulted in ultrasound systems commercialized by the spin-out company us4us Ltd. [103]. One of these devices is the us4R-lite system — a low-cost, portable ultrasound research platform with a reduced receive channel number and an optional embedded low-power GPU module for processing acceleration. Although it has limited hardware resources, this system is suitable for the implementation of advanced ultra-

Table 1.2: Summary of commercially available systems implementing real-time 2-D SWE method. Adopted from [79].

Manufacturer	Device	Method	Manufacturer name
Supersonic Imagine (Aix en Provence, France)	Aixplorer, Aixplorer Ultimate, Aixplorer MACH 20 & 30	SSI	ShearWave Elastography
Siemens (Munich, Germany)	Acuson S2000 & S3000	SWEI	Virtual Touch Quantification (VTq)
Philips (Amsterdam, Netherlands)	EPIQ 5, EPIQ 7	SWEI	ElastQ
General Electric (Boston, MA, USA)	Logiq E9, E10, E10s	CUSE	Shear Wave Elastography
Canon Medical Systems (Otawara, Japan)	Aplio 300 & 400 & 500	SWEI	Shear Wave Elastography
Mindray (Shenzhen, China)	Resona 7, DC-80	SWEI	Sound Touch Elastography
Verasonics (Kirkland, WA, USA)	Vantage	-	Open Research System
S-Sharp (New Taipei City, China)	Prospect T1	-	Open Research System

sound imaging modes and algorithms. As a research platform with high configurability of acquisition schemes, it also allowed to perform research and optimization within a given imaging mode. The author of this thesis as an employee of us4us company was familiar with the low-level architecture of the us4R-lite system to which he contributed to its digital part design. In addition, the author gained a lot of experience in system power optimization while working as a consultant for an American startup company developing a state-of-the-art handheld point-of-care ultrasound scanner utilizing cutting-edge pMUT technology [104].

1.4 Thesis aims

Summing up all the reasoning above, the following aims were defined for this thesis:

- implement the two-dimensional (2-D) shear wave elastography (SWE) using the portable us4R-lite ultrasound research platform;

- develop data processing of the performance providing real-time 2-D SWE imaging capabilities;

- evaluate and optimize the 2-D SWE technique implementation to make it more suitable for selected point-of-care class devices.

The expected practical results of the efforts above would fit well the idea behind the Ph.D. program in cooperation with an industry partner (*"Doktoraty wdrożeniowe*[1]*"*)

---

[1]Doktoraty wdrożeniowe is a program held by Polish Ministry of Science and Higher Education. More details on that program can be found on the Polish Government's web-

which the author is a participant of. The specific subject of this thesis can be formulated as follows.

> **Thesis:**
> The real-time 2-D shear wave elastography imaging can be practically implemented using a point-of-care device class by exploiting a GPU-accelerated, software-defined ultrasound approach.

In the thesis — as formulated above — term *2-D shear wave elastography* is understood as any ultrasound-based technique that allows the reconstruction of quantitative two-dimensional elasticity maps by estimation of shear wave propagation speed; term *implementation* is understood as designing the hardware/software framework able to perform data acquisition and reconstruction of elasticity maps; term *point-of-care device class* means portable devices of dimensions much smaller than typical cart-based systems, that contain all the components needed to perform ultrasound acquisition and are optimized in terms of power consumption. Operation on a battery is desirable, but not required; term *software-defined ultrasound* refers to ultrasound imaging systems design paradigm in which hardware processing is limited to favor data processing routines through computer programming. In such systems, hardware processing is often limited to collecting data, whereas image reconstruction or other processing is performed entirely in software. Finally, the term *real-time* is understood as *providing instantaneous feedback* and featuring an image refresh rate of a minimum 1 frame per second.

Implementation and experimental works held in order to reach aims and prove the thesis as defined above are described in 5 chapters of this manuscript. The contents of each chapter are described briefly in the next section.

1.5 Thesis outline

The structure of the thesis in the form of a block diagram is presented in Fig. 1.8. The thesis is organized as follows:

- **Chapter 1** (this chapter) provided background information showing how the developments in the field of electronics dictated the advances in ultrasound imaging system design that — among others — made the implementation of tissue elasticity imaging feasible. Next, the theoretical basis of ultrasonic tissue elasticity imaging was formulated. Then, a review of relevant methods developed through the years was done. Identification of trends and limitations observed in state-of-art ultrasound imaging and elastography methods particularly, allowed to define aims to be achieved by this thesis.

- **Chapter 2** presents the theoretical background of the 2-D shear wave elastography method. Various aspects of this technique are reviewed — including known approaches of spatial modulation techniques used in pushing beam generation, data acquisition schemes, and methods used in data processing for elasticity maps
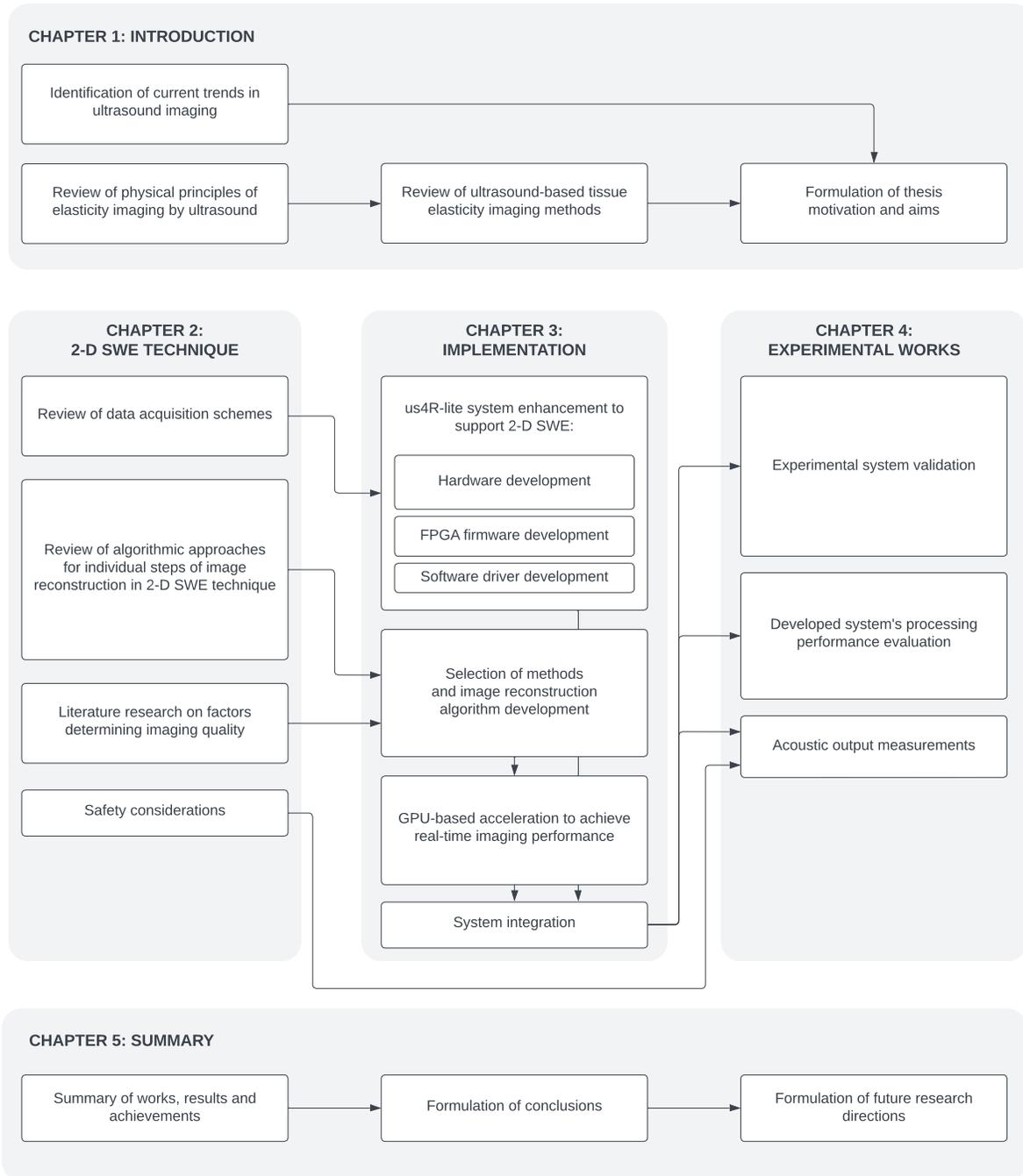
Figure 1.8: A conceptual organization of the thesis. Blocks represent the core tasks within each of the chapters. The size of each block illustrates coarsely the relative amount of effort put into each of the tasks. The arrows in the diagram show the relationships between theoretical, practical, and experimental parts of the thesis.

reconstruction. In each of those fields, along with commonly used methods, novel approaches are identified. After that, safety issues are covered by considering the limitations aimed to avoid undesired tissue bioeffects during SWE imaging procedure. The remaining part of the chapter discusses sources of uncertainty and bias in 2-D SWE method. Furthermore, the parameters affecting the spatial resolution of the images obtained in SWE are considered.

- **Chapter 3** describes the implementation of the 2-D SWE technique on the us4R-lite ultrasound research platform. Firstly, the hardware part is covered — a top-level architecture of the us4R-lite system and the development of a dedicated transmit extension board enhancing systems transmit generation capabilities suitable for SWE. Then, an overview of FPGA firmware architecture and functionality is provided. Next, a description of the software implementation of the image reconstruction algorithm is given. The remaining part of the chapter focuses on efforts targeted to achieve real-time imaging. The problem is addressed firstly by estimation of the required computational performance and data throughput of the system, which allowed to make proper decisions on the software architecture level. Since the implementation method assumed the usage of GPU, a brief overview of general-purpose programming of massively parallel processing using GPU is provided. Topics covered include: CUDA-capable GPU architecture, parallel processing threads execution, and software scalability. Then, the considerations on 2-D SWE image reconstruction problem suitability to exploit parallel processing are followed by a detailed description of processing kernels related to each reconstruction algorithm's step. Along with numerical problem decomposition for parallel processing and a general description of data processing implemented by each of the kernels, the optimization techniques that were utilized to enhance the processing performance are discussed. The chapter concludes with a summary of data processing partitioning between software and hardware parts of the system.

- **Chapter 4** contains experimental part of the thesis. It consists of three main parts. In the first part, the acoustic output of the system was measured to evaluate the safety of the imaging procedure. The second part presents the results of the developed system validation and 2-D SWE imaging quality assessment. The operation of the system was verified experimentally, and the system performance of the shear wave speed measurement was evaluated by deriving chosen quality metrics of images acquired using industry-standard elastography phantom. Moreover, some intermediate results are shown to expose some typical problems appearing in SWE mode that may affect the imaging quality. In the third part, the developed system's processing performance is evaluated and the impact of chosen reconstruction parameters on processing time and imaging quality is evaluated.

- **Chapter 5** concludes this thesis by providing a summary of the works presented. Then, after the thesis's main achievements are described, limitations are discussed, possible future research directions are proposed and the final conclusions are formulated.

Chapter 2

Shear wave elastography technique

The previous chapter introduced the basic physical fundamentals of ultrasound-based elasticity imaging methods based on shear wave propagation speed estimation. Several methodological approaches were also briefly discussed. This chapter discusses in detail one method, the 2-D shear wave elastography technique, that was chosen to be implemented. In this method, an impulsive ARF is used as an excitation source and resulting shear wave tracking is performed off the axis of the push excitation beam in order to reconstruct 2-D tissue elasticity maps. The same transducer is used to both generate the ARF and track the shear waves propagation. Hereinafter in this thesis, this technique will be named shortly under the abbreviation 2-D SWE or simply SWE.
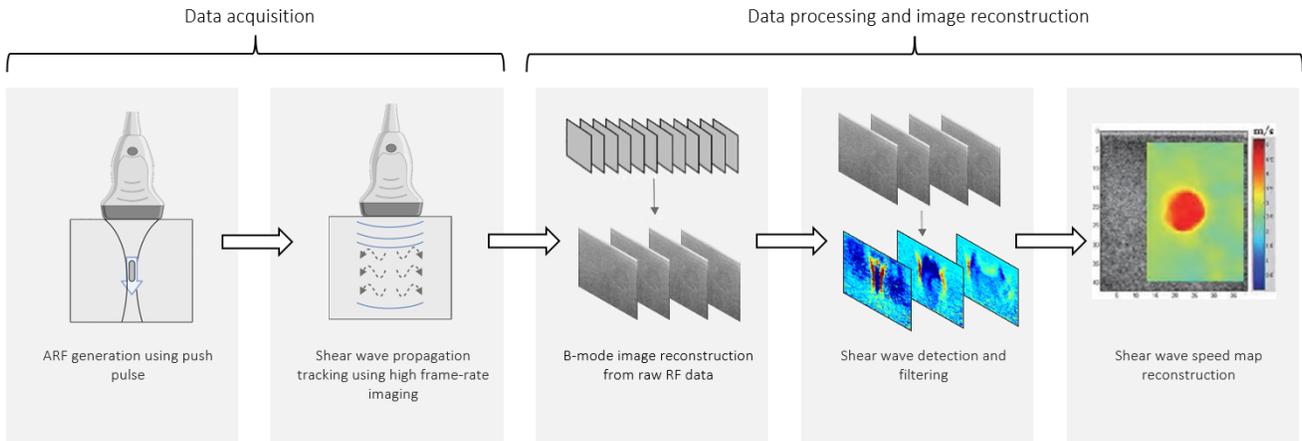
Figure 2.1: A general measurement procedure in the 2-D SWE technique.

A general multi-step SWE measurement procedure is shown in Fig. 2.1. Same as the vast majority of other ultrasonic techniques, the whole process consists of two stages: data acquisition followed by data processing with image reconstruction. Firstly, the ARF is generated in the medium using a push pulse. The shear waves motion resulting from the impulsive excitation of ARF is then tracked ultrasonically by detecting small axial displacements along the shear wave propagation path using pulse-echo ultrasound. There are various acquisition strategies possible to perform shear wave propagation tracking. In one commonly used method, immediately after the end of push generation, a scanner switches into high frame-rate imaging mode and captures RF data. RF data

are then used in the beamforming process to reconstruct a series of images. A series of beamformed images are the input of the correlation-based algorithm that performs shear wave detection by obtaining particle axial displacement versus time at each pixel position. Finally, after spatio-temporal filtering, shear wave motion data is used to obtain a local shear wave speed (SWS) map which reflects local tissue stiffness. Each of these steps is discussed in detail in the following sections — commonly used techniques are described and the key trade-offs are identified.

2.1 Push generation

Push generation is the first step in the SWE technique procedure and it dictates conditions for all following steps defining largely the final image quality. The ultrasound beams used for ARF generation (pushing beams) in the SWE method are generated by an array transducer. In the same way as in conventional imaging, electronic transmit beam focusing and steering can be done by applying a proper time delay profile and apodization (optionally) over the active aperture. The pushing beam induces shear waves within a two-dimensional $(x - z)$ field-of-view (FOV) in which shear wave energy distribution is nonuniform and is strongly dependent on both tissue properties and ultrasonic beam characteristics. It is desired to distribute shear waves of detectable axial displacement amplitudes within the FOV to provide good conditions for shear wave detection in as large region as possible. Therefore, since the very beginning of ARF-based SWE methods, there have been a lot of research efforts put into a single pushing beam optimization and also an exploration of various sequences of pushing beams to generate desired shear wave motion distribution.

A) Single pushing beam parameters

The amplitude of the shear wave-induced displacements increases along with deposited ARF intensity. Therefore, for bigger displacement amplitudes, one can increase the transducer electrical excitation voltage, push duration, or both for improved signal-to-noise ratio (SNR) of shear wave detection and, as a result, a higher likelihood of successful SWS estimation [105]. Excitation voltage in clinical systems is typically in a range of 100–180 Vpp. Push duration is typically in a range of 0.1–1 ms, depending on the system transmit capabilities and power plan. However, disposed energy shall not exceed allowable thresholds. See section 2.4 for details.

Normally, the push transmit frequency is set to the center frequency of the transducer to maximize transmit efficiency. However, there are often two reasons to consider using a lower push frequency (but still within the bandwidth of the transducer). Firstly, employing a lower push frequency helps to broaden the push beam width to reduce the underestimation of tracked tissue displacement due to speckle shearing within the track point spread function (PSF) and improves shear wave displacements jitter upon detection [106], [107]. Secondly, a lower push frequency can enhance push pulse penetration depth due to reduced attenuation at lower frequencies. However, as indicated by Eq. (2.11), for the same pressure amplitude, lower pushing frequencies will cause the MI limit to be encountered sooner. Hence, the excitation frequency can be also adapted for specific focal depth, medium attenuation, and transducer bandwidth to improve energy transfer efficiency and obtain larger displacements while being still within the FDA limits [108].

The aperture size used for push pulse generation is typically chosen along with focal depth. Those two parameters define the f-number of the beam: $F_\# = z_f/D$, where z_f is the focal depth and D is the aperture size. The f-number controls a trade-off between shear waves' maximum displacement amplitudes and axial shear wave field size. Depending on the f-number of the pushing beam, the resulting motion can be very confined in the axial direction (for low f-numbers) or can be very broad (for high f-numbers). As a result, using more aggressive focal configurations to generate higher peak displacement amplitudes can be done, but at a cost of reduced shear wave field size within the FOV.

B) Pushing sequences — spatial modulation strategies
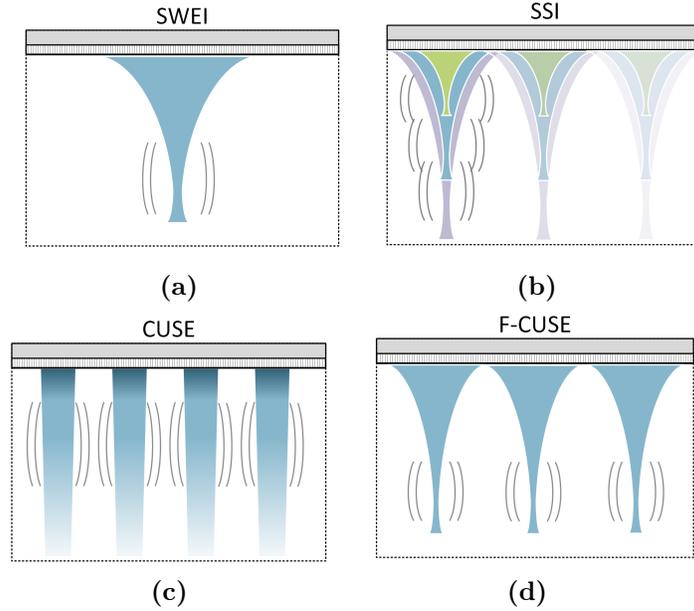
Figure 2.2: Push generation diagrams for commonly used techniques. (a) SWEI — a single focused beam. (b) SSI — a sequence of beams focused at successive depths. Three acquisitions are done with 3 different pushing beams' lateral positions. (c) CUSE — an unfocused comb-push. (d) F-CUSE — focused comb-push.

Shear wave attenuation is high in soft tissues, so the shear waves do not typically travel very far (0.5–2.0 cm) [79], depending on the shear wave energy. To make an image that covers the whole FOV, a sequence of spatially distributed pushing excitations is necessary. The shear wave velocity results from each acquisition are then combined to form the final compounded image. The researchers used various sequence designs and some commonly used methods are presented in Fig. 2.2. In early works, Nightingale et al. used a single, focused push beam to generate an ARF [59]. However, as a result of focusing, the shear waves were generated in close axial proximity to the focal point of the pushing beam. Therefore, high-quality elasticity map reconstruction was possible only in limited depth-of-field (DoF). To address this issue, a new push beam generation method was proposed by Bercoff et al. — Super Sonic Shear Imaging (SSI) [61]. In SSI, a sequence of beams focused at consecutive depths was used to generate a broad shear wavefront over the high axial extent to cover a big part of FOV. Although even a single SSI acquisition covers a big part of FOV, there is still a lack of shear waves on

the beam axis. Therefore, the authors used 3 acquisitions, each with a different pushing beams' lateral position [63]. An alternative approach, named Comb-push Ultrasound Elastography (CUSE), was proposed by another group in a series of works ([109]–[112]), where multiple beams arranged in a comb pattern were used to generate shear waves. Adjacent push beams produced shear waves that covered other push beams' axes. As the comb-push beam induces shear waves in the entire FOV (laterally), an elasticity map can be reconstructed in the whole FOV by only one acquisition. They tried to use both unfocused and focused beams organized in a comb. Nevertheless, unfocused beams forming a comb-push featured low penetration in depth, while focused beams had limited DOF. In [113], the authors used a hybrid beamforming approach to optimize a single pushing beam shape. They combined two types of focusing, conventional spherical focusing, and axicon focusing, to produce a beam for generating a shear wave that has increased DoF. Recent developments show, that even more complex beam patterns, like Bessel-apodized beams [114], can be used to achieve more optimized shear wave energy distribution within the FOV and improved power efficiency.

2.2 Data acquisition sequences for shear waves tracking

After the push generation, shear waves induced in the tissue must be observed in order to deduce the mechanical properties of the medium from its propagation. Shear wave speed (SWS) depends on the characteristics of the medium and it generally varies between 1 and 20 m/s in biological tissues. For instance, for a medium that has a Young modulus of 50 kPa, using Eq. (1.19) one can obtain SWS of $\sim$4.1 m/s, which means, that shear wave propagates laterally through the example 40 mm wide field-of-view (FOV) of the probe in $\sim$9.75 ms. It is therefore a very fast phenomenon that must be imaged with a high temporal resolution. In conventional ultrasonic scanners that acquire one scan line at a time employing transmit focusing, it takes approximately several tens of milliseconds to produce one image frame, which is far too slow for shear wave tracking.

There are two common approaches to achieving sufficient shear waves tracking temporal resolution. In the first solution used in early works (e.g. [59]), standard focusing is used for the generation of parallel tracking beams at several lateral locations distal to a shear wave source. The acquisition sequence is then designed to quantify displacement through time within a ROI. The push pulse is generated repeatedly at the same location and each push pulse is followed by a single or several interleaved tracking lines to observe the shear wave propagation. Then, the location of the tracking lines is sequenced across the ROI. The lateral spacing of the tracking lines is determined by the desired ROI size. This method allows the utilization of dynamic receive focusing to improve signal SNR. This method is suitable for conventional scanners, however, an important drawback of this technique is that the push pulse must be generated multiple times, which increases the total data acquisition time, limits the imaging frame rate, and is sub-optimal in terms of power efficiency.

In the second approach — Plane Wave Imaging (PWI) — instead of employing the transmit focusing, a plane wave is transmitted, and the echo data are acquired at all receive elements. This whole dataset is then used in software to perform parallel receive beamforming. Since the plane wave insonifies a whole FOV, the entire image can be produced from a single transmit-receive event. Without using transmit focusing, this method can achieve frame rates as high as 10,000 fps, depending on the imaging depth,

which allows shear wave tracking with high temporal resolution. The typical acquisition scheme used in this method is shown in Fig. 2.3. Obviously, due to the lack of transmit focusing, PWI is less effective for contrast and lateral resolution. Nevertheless, in most cases, since the generated shear wavelengths are much greater than the resolution of the ultrasound image obtained, the loss of resolution does not significantly influence the quality of the shear wave motion estimation [115]. In PWI, the image resolution can be enhanced by transmitting plane waves in different directions and compounding the image from data obtained from every angle, which was proposed in seminal work by Montaldo et al. [116] — Coherent Plane Wave Compounding (CPWC). CPWC reduces the frame rate by the number of angles but remains fast enough to follow transient shear wave motions. The more angles are used, the higher improvement is observed in contrast and resolution, but the frame rate decreases accordingly. For efficient shear wave tracking both high temporal resolution and high image quality are desired. CPWC can achieve image quality comparable to the multifocal imaging method, but it requires tens of angled transmissions. Consequently, a trade-off between frame rate and image quality is chosen depending on the application. To preserve the frame rate with angle compounding, a sliding window with a moving average along the slow time can be utilized. Recently, the time-aligned plane-wave compounding method was introduced for shear wave tracking by Capriotti et al. to maintain a high PRF [117]. This method employed interpolation to obtain data points at higher frame rates, and the time-aligned data were compounded to increase the SNR. Other modifications that involved advanced transmit pulses generation to improve PWI data for shear wave tracking were also reported. Papadacci et al. investigated using diverging waves instead of plane waves [118]. Independently, Doherty et al. [119] for the ARFI method and Song et al. [120] for 2-D SWE, proposed pulse-inversion harmonic imaging for improvement of displacements measurement accuracy in the presence of severe noise sources such as phase aberration, ultrasound reverberation, and clutter noise. The same group also investigated the feasibility of using coded excitation in PWI for shear wave detection, to increase the shear wave motion SNR, especially at greater depths [121].
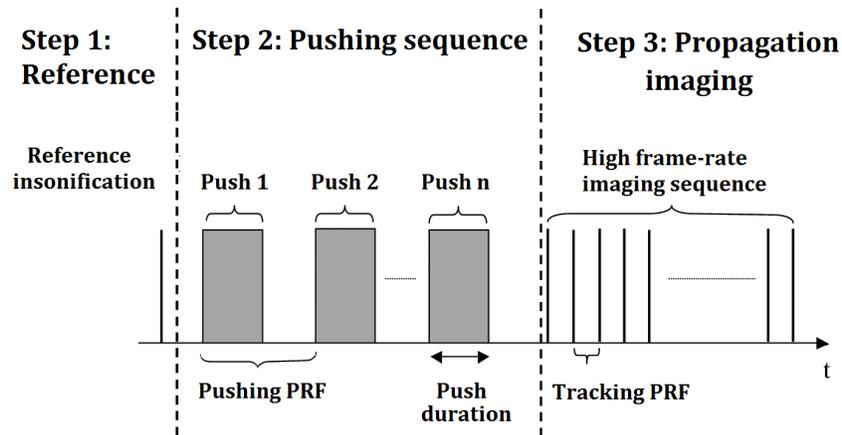
Figure 2.3: A general acquisition scheme for the SWE technique using high frame-rate PWI for shear waves tracking. In the first step, an optional reference image is captured. Then, a push pulse or a sequence of push pulses is generated. Immediately after that, the scanner switches into high frame-rate imaging mode and captures RF data. Some important parameters of acquisition are shown in the image: pushing PRF, tracking PRF, and push duration.

2.3 Data processing and image reconstruction

In the 2-D SWE method, raw RF channel data obtained during acquisition are used as input to the reconstruction algorithm which typically consists of multiple steps:

- pre-processing and receive beamforming,

- shear wave detection,

- shear wave motion data filtering,

- shear wave speed estimation,

- image compounding, and

- post-processing.

The most relevant methods used in each of these steps are described in the following sections. Most attention is paid to methods that were implemented practically later in this thesis. Novel approaches are also discussed.

2.3.1 Pre-processing and receive beamforming

In the first step of the algorithm, the RF channel data is band-filtered to filter out frequency components out of a transducer band coming from electronic noise and improve this way SNR of the RF data. Then, there can be various beamforming strategies used, depending on the acquisition scheme. Most commonly, a classical DAS (delay-and-sum) beamforming is performed on PWI data. Moreover, numerous beamforming techniques have been proposed in recent years to compensate for reduced PWI image quality, both in terms of resolution and contrast and this problem was a subject of challenge during the 2016 *IEEE International Ultrasonics Symposium* in Tours, France [122]. However, plane wave beamforming is not specific to 2-D SWE method and a review of beamforming techniques is out of the scope of this dissertation. The curious reader can check other sources for reference, e.g. [116].

The pre-processed input data is the 3-D dataset $RF(n, m, T)$, where n and m are sample number and receive channel number, respectively, and T is the number of frame, representing the slow-time domain. For the purpose of further analysis, it is assumed, that the beamforming process takes this input and generates a complex 3-D dataset: $R(x, z, T) = I(x, z, T) + jQ(x, z, T)$, where I and Q are the in-phase and quadrature components of the beamformed and demodulated signal, x and z are lateral and axial dimension indices, respectively. I/Q data can be obtained using quadrature demodulation or Hilbert transform. Envelope detection and compression are not performed, since image data lacks the phase information needed to track micron scale displacements typical in the SWE technique [123]. However, B-mode images are used for guidance in real-time applications.

2.3.2 Shear waves detection

Shear wave motion produces small tissue displacements that result in a difference between successively acquired ultrasound frames. There are many time-delay estimators available for motion tracking. In general, they can be classified into two categories:

phase-shift algorithms that operate on I/Q data and correlation-based methods that operate on RF data.

A) Phase-shift methods

In the case of phase-shift algorithms that are derived for SWE from Doppler methods, axial displacements of shear waves are apparent as a phase difference $\delta\varphi$ in the corresponding pixels of consecutive beamformed frames. Detection of this phase shift with respect to the center frequency allows to obtain a local axial velocity v_z:

$$v_z(x,z) = \frac{c \cdot \delta\varphi(x,z)}{4\pi f_c \cdot FRI}, \tag{2.1}$$

where c is the speed of sound (average $1540\,\mathrm{m/s}$ in soft tissues), f_c is the tracking pulse center frequency, and FRI is the frame repetition interval. Removing FRI from the Eq. (2.1) gives the local axial displacement $u(x,z)$. The Kasai 1D-autocorrelator algorithm [124] is commonly used to calculate $\delta\varphi$ for each pixel of each frame (Eq. (2.2)). This method measures the average phase shift over an axial sample range M and N samples in the slow time domain (ensemble length) to obtain 3-D shear wave motion data $v_z(x,z,T)$. M and N can be adjusted in a trade-off between displacement estimation accuracy and spatio-temporal resolution of motion data. In the formulation of (Eq. (2.2)), the signal at each time step is compared with the previous time step, which is called progressive referencing. Alternatively, the reference frame acquired before the push pulse (see Fig. 2.3) can be used as the reference to calculate the displacement at each time step. This approach is referred to as a fixed reference scheme.

$$\delta\varphi = \arctan\left(\frac{\sum_{n=0}^{N-2}\left[\sum_{m=0}^{M-1} Q(m,n)\sum_{m=0}^{M-1} I(m,n+1) - \sum_{m=0}^{M-1} I(m,n)\sum_{m=0}^{M-1} Q(m,n+1)\right]}{\sum_{n-0}^{N-2}\left[\sum_{m=0}^{M-1} I(m,n)\sum_{m=0}^{M-1} I(m,n+1) + \sum_{m=0}^{M-1} Q(m,n)\sum_{m=0}^{M-1} Q(m,n+1)\right]}\right) \tag{2.2}$$

Loupas' 2-D autocorrelator [125] is another commonly used phased-shift algorithm for shear wave motion detection. It can be written as in Eq. (2.3). Whereas in Kasai's algorithm it was assumed that the mean frequency of the signal f_c is constant, Loupas' algorithm enhances displacement estimation by calculating the mean Doppler frequency and the mean signal frequency along each axial extent. Kasai's assumption of a fixed center frequency is inaccurate for small ranges of the RF signal, therefore, such correction for the local variation in center frequency yields more accurate estimates [126]. Phase-shift algorithms only estimate displacements u in the range $-\lambda/2 \leq u \leq \lambda/2$. Lee et al. in [127] derived theoretically and experimentally that using a phase-shift algorithm on frames that come from CPWC can suffer from an inherent displacement estimation error since the absolute value of the phase difference can be larger than $\lambda/2$ at some pixels. They proposed a simple solution to eliminate that problem, named initial-phase-compensated plane wave compounding (IPCPWC), which allows achieving increased

SNR and reduced jitter of shear wave motion data.

$$v_z = \frac{c}{4\pi \cdot FRI \cdot f_c} \cdot \frac{\arctan\left(\dfrac{\sum\limits_{m=0}^{M-1}\sum\limits_{n=0}^{N-2}[Q(m,n)I(m,n+1)-I(m,n)Q(m,n+1)]}{\sum\limits_{m=0}^{M-1}\sum\limits_{n=0}^{N-2}[I(m,n)I(m,n+1)-Q(m,n)Q(m,n+1)]}\right)}{1+\frac{1}{2\pi}\arctan\left(\dfrac{\sum\limits_{m=0}^{M-2}\sum\limits_{n=0}^{N-1}[Q(m,n)I(m+1,n)-I(m,n)Q(m+1,n)]}{\sum\limits_{m=0}^{M-2}\sum\limits_{n=0}^{N-1}[I(m,n)I(m+1,n)-Q(m,n)Q(m+1,n)]}\right)} \tag{2.3}$$

B) Correlation-based methods

The development of precise and robust time delay estimators for ultrasound has been the focus of research for many years and new methods still have been proposed. Viola et al. compared the performance of various estimators, showing that sum squared differences, normalized correlation, and normalized covariance outperform the other algorithms by most measures relevant to time-delay estimation in ultrasonic signals [128]. In particular, a normalized cross-correlation (NCC) is often used for rapid displacement tracking. The normalized cross-correlation function, $c(f_r, f_d, j)$, between the reference, f_r, and delayed, f_d, signals at a lag j is normalized by their standard deviation [126]:

$$c(f_r, f_d, j) = \frac{\sum\limits_{i=-M/2}^{M/2}[f_r(i) - \bar{f}_r] \cdot [f_d(i+j) - \bar{f}_d(j)]}{\sqrt{\sum\limits_{i=-M/2}^{M/2}[f_r(i) - \bar{f}_r]^2 \cdot \sum\limits_{i=-M/2}^{M/2}[f_d(i+j) - \bar{f}_d(j)]^2}}, \tag{2.4}$$

where $\bar{f}_r$ is the mean of the reference signal over a window of size M and $\bar{f}_d(j)$ is the mean of the delayed signal over a window shifted by j samples. The argument of the maximum of the NCC function indicates the point at which two signals are most similar to each other, and this measure can be used to determine the displacement between a reference and a delayed signal. In NCC, timing precision is defined by the system sampling frequency. With a sampling frequency of 65 MHz, an axial displacement of one sample is approximately 12 μm, which is generally larger than shear wave axial displacements in SWE technique. Therefore, it is necessary to upsample the data before NCC to provide the ability to resolve small displacements. Various interpolation methods can be used for that purpose, i.e., Pinton et al. [126] combined cubic spline interpolation of the RF samples and parabolic fitting to the maximum of the correlation function. Correlation-based methods by their nature and due to the requirement of upsampling are more computationally intensive than phase-shift algorithms so the latter are more readily used in real-time applications [123].

C) Novel approaches

Generally, in SWE, due to the requirement of a high frame rate and various sources of noise, the input data for shear wave motion detection often suffer poor quality, which results in the low SNR of shear wave motion data. This issue is especially significant at bigger depths and distant from push ROE. Therefore, to address this issue, in addition to advanced acquisition techniques as discussed in the section 2.2, researchers keep proposing novel algorithms for shear wave detection to improve motion data quality. In

[129] and later in [130] authors proposed a complex method incorporating knowledge of adjacent displacement estimates into the final estimate of a displacement utilizing biased estimation using Bayes' theorem. Their results have shown, that by the cost of slightly increased bias, a significant drop of estimates variance can be achieved. However, the proposed algorithm featured high computational complexity. Horeh et al. formulated tissue displacement estimation as an optimization problem and proposed two computationally efficient algorithms to estimate the shear wave displacement field achieving promising results [131].

2.3.3 Shear wave motion data filtering

After shear wave motion data is obtained, several filtering steps can be performed to reduce the noise contents and to enhance data SNR before SWS estimation. These steps include band-pass filtering and directional filtering in the Fourier (k_x, ω) space. Both steps are discussed below.

A)   Frequency filtering

Signal decorrelation, finite window lengths, and noise can result in random errors in the tissue motion data [123]. In order to attenuate high-frequency jitter, a low-pass filter can be applied to the displacement profile at each pixel location in a slow time direction. The cut-off frequency of the filter has to be chosen high enough so that shear wave energy is not rejected by the filter. Since the spectral content of the shear waves depends not only on excitation parameters but also on the mechanical properties of the tissue, it is not known in advance. For this reason, it is reasonable to pick the cut-off frequency according to shear wave energy spectral analysis.

For *in vivo* applications, other problems such as motion artifacts should be addressed. In such cases, motion filters are commonly applied to reduce the motion artifacts arising from biological motion or transducer motion. A high pass filter can serve as a motion filter because anatomical motion is at a low frequency ($<$10 Hz), while the shear waves typically have a center frequency greater than 50 Hz [123], and the low-frequency motion can be easily filtered. For SWE cases, where the high-pass filter fails, Mellema et al. proposed utilizing the Empirical Mode Decomposition (EMD) as a parameter-free approach to decouple shear wave motion data from other motion artifacts [132].

B)   Directional filtering

When shear waves propagating through the medium encounter the interface where there is a change in stiffness, reflections occur. Such reflected waves at the lesions or stiffer inclusion boundaries can result in artifacts in SWS estimates [133] because the incident and reflected waves propagate in opposite directions and are present simultaneously. Therefore, to address this problem, a directional filter is used to reduce the shear wave reflections. Based on the spatio-temporal directional filter proposed by Manduca et al. in [134] for Magnetic Resonance Elastography (MRE), a dedicated algorithmic approach was described by Deffieux et al. [135]. In this work, the authors proposed using a 2-D filter in the Fourier (k_x, ω) space. The motion data was sliced along depth dimension (z) to form two-dimensional slices $v_z(x, T)$. A 2-D Fast Fourier Transform (FFT) can be used to transfer each slice of data to the Fourier domain. Then, filtering is performed by simple multiplication by masks designed to extract shear waves propagating left-to-right (L-R) — v_{zLR} — and right-to-left (R-L) — v_{zRL} and the results are brought back to the

spatio-temporal domain using inverse 2-D FFT. As a result, two 3-D shear wave motion datasets are constructed. This process is diagrammed in the Fig. 2.4.
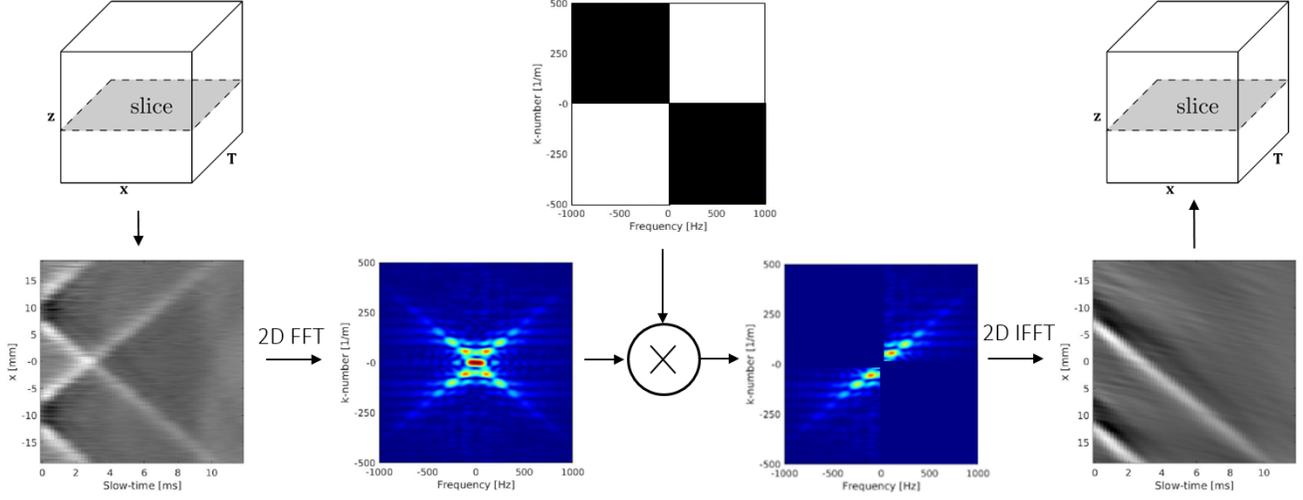
Figure 2.4: Principle of directional filtering. The 3D motion data $v_z(x, z, T)$ is sliced along depth dimension (z) to form 2D slices $v_z(x, T)$. Then, a slice of spatio-temporal data is transferred to the (k_x, ω) space using 2-D FFT. This domain representation can be separated in four quadrants, $k > 0$ and $\omega > 0$, $k < 0$ and $\omega < 0$, $k < 0$ and $\omega > 0$, and finally $k > 0$ and $\omega < 0$. Depending on the sign of the phase velocity $c = \omega/k$, forward $(c > 0)$ and backward $(c < 0)$ propagating components can be separated by simple multiply in the (k_x, ω) space by a mask. After this operation, an inverse 2-D FFT is used to transfer data back to the spatio-temporal domain.

The usage of a directional filter is especially important in push generation techniques such as CUSE, where shear waves from multiple sources overlap [109]. It should be also noted, that a directional filter improves the shear wave motion data SNR. By removing nearly half of the (k_x, ω) domain in which nearly no useful signal is present but where half of the white noise energy is distributed, the filter can improve the SNR by up to 2-fold [135]. In addition, masks can be designed in a way to filter out all data components related to shear waves propagating speeds out of range of interest. The mask edges should be also apodized to avoid sharp cuts in the spectrum and subsequent Gibbs oscillations in the spatio-temporal domain [135]. Song et al. described a similar approach in the CUSE technique to decompose the complex shear wave field into differently oriented shear waves for compounding purposes [111]. Higher dimensional directional filters were also proposed to deal with in-plane and out-of-plane interfaces [136], [137].

2.3.4   Shear wave speed estimation

A)   Direct Helmholtz equation inversion

To estimate the stiffness of tissue, a 2-D SWS map $v_{SW}(x, z)$ can be estimated from 3-D displacement data. The obvious approach used in early SWE works relied on computing SWS as the algebraic inversion of the Helmholtz equation (Eq. (1.16)), which relates SWS and displacement. While Nightingale et al. used pseudo-direct inversion of the Helmholtz equation [59], Bercoff et al. in his seminal paper introducing the SSI method proposed the Helmholtz inversion in the Fourier domain [61]. This reconstruction method requires a calculation of second-order spatial and temporal derivatives.

Nonetheless, differentiation of motion data which is typically jittered and have poor SNR results in inaccurate SWS values.

B) Time-of-flight estimation methods

The above challenges have motivated the use of time-of-flight (ToF) methods for SWS estimation. ToF methods take advantage of *a priori* information about the shear wave propagation direction — in 2-D SWE, the shear waves are assumed to travel in the lateral direction. In general, ToF methods can be classified into three groups: linear regression, correlation-based, and shear wave trajectory detection methods.

Linear regression methods first determine the shear wave arrival time at each lateral position and then calculate the slope of the position versus time data. The slope is calculated using data from positions within a lateral kernel that is stepped across the ROI. To obtain slope Deng et al. used least square minimization [107], while Wang et al. proposed the application of the Random Sample Consensus (RANSAC) algorithm [138], which appeared to be more robust to the presence of gross outlier data. Kernel sizes range from 0.5 to 6 mm [139]. In ToF methods motion profiles need to be interpolated before shear wave arrival calculation to achieve fine temporal resolution. The arrival time can be estimated directly from the shear wave motion data using chosen features of the signal, like time of peak displacement [60], [139] or time of peak particle velocity [139].

Another common method estimates the local SWS using cross-correlation of tissue motion signals at adjacent spatial locations [63], [109], [110], [139], [140]. In this method, a normalized cross-correlation (NCC) as defined in Eq. (2.4) is calculated of shear wave axial velocity profiles along the lateral dimension m and separated by lateral kernel size Δd but on the same depth n to find the shear wave arrival delay Δt:

$$\Delta t(x, z) = \frac{1}{FRI} \cdot \arg \max_j c\left(f_r(x, z), f_d(x, z), j\right), \tag{2.5}$$

where $f_r(x, z)$ and $f_d(x, z)$ are respectively:

$$f_r(x, z) = v_z(x - \Delta d/2, z, T) \tag{2.6}$$

$$f_d(x, z) = v_z(x + \Delta d/2, z, T) \tag{2.7}$$

To facilitate more accurate cross-correlation results, a window function can be applied to the shear wave motion profiles. This calculation was repeated to get Δt and cross-correlation coefficient pairs for each image pixel within the ROI. Local SWS v_{SW} in each pixel can then be obtained by a simple formula:

$$v_{SW}(x, z) = \frac{\Delta d}{\Delta t(x, z)} \tag{2.8}$$

This method is diagrammed in the Fig. 2.5. Song et al. increased the robustness of the correlation-based method by addressing the estimation of SWS from multiple differently angled push pulses [111]. In this 2-D approach, SWS along both the lateral and axial directions were estimated, and the vector sum of the two speeds was calculated. To increase estimation accuracy, the final result was a weighted sum of multiple SWS estimations along the lateral and axial directions.

Another ToF-based technique of SWS estimation was proposed by McLaughlin and Renzi in [141], where they proposed a family of methods, to solve the inverse Eikonal
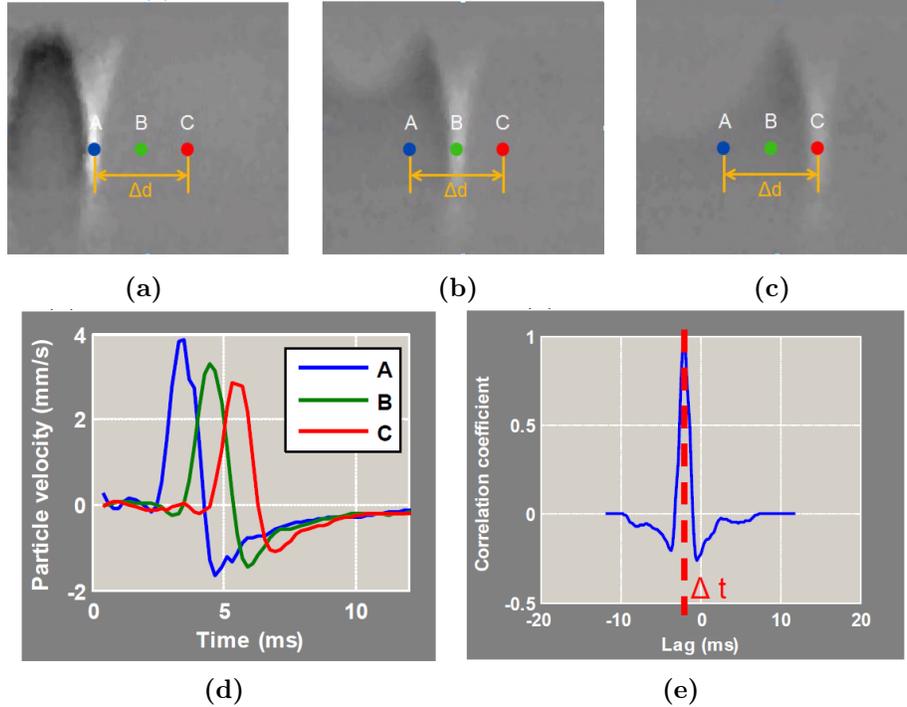
Figure 2.5: Principle of correlation-based local SWS estimation. (a)–(c): shear wave motion frames of a single shear wave propagation through the ROI. Three pixels are chosen. Pixel B is the targeting pixel where local SWS estimate is desired. Shear wave signals at neighboring pixels A and C within the lateral kernel Δd are used for estimation (d): Axial particle velocity profiles at chosen pixels. (e): Lag for max NCC value of profiles A and C gives the estimate of time delay, which can be used to obtain SWS with a known distance between pixel A and pixel C, which is Δd. Image source: [133].

equation. However, these methods were computationally intensive. Recently, Lee et al. [142] proposed a robust and computationally efficient 2-D SWS calculation, named the Tangent Plane Method (TPM), which combined the least square method to solve the Eikonal equation with 2-D correlation-based approach similar to [111].

C) Shear wave trajectory methods

Shear wave trajectory methods of SWS estimation rely on the identification shear wave trajectory from the spatio-temporal shear wave motion data (see Fig. 2.4 for example). In this approach, SWS is equal to the slope of the spatio-temporal trajectory with the maximum energy. Rouze et al. used the Radon sum transformation to find the shear wave trajectory [143], i.e. the optimal trajectory was identified by the peak Radon sum. Carrascal et al. improved the precision of this algorithm by introducing two methods for enhancing SWS estimation — the first method combined spatial and temporal peak searches in shear wave motion data to estimate shear wave velocities, whereas the second method employed spatio-temporal thresholding to estimate shear wave velocities [144]. These methods however are useful only for a homogeneous medium. Recently, Li et al. proposed a real-time curve tracing (RTCT) method for shear wave trajectory detection, that was suitable in both homogeneous and heterogeneous cases and showed good performance in the presence of noisy motion data [145].

2.3.5 Image compounding and post-processing

The local SWS estimate can be obtained wherever the shear wave strong enough has propagated. Depending on the push generation design, there can be regions where SWS estimation fails due to a lack of the shear wave or if the shear wave energy propagating in the estimation point and its proximity was too low in the presence of noise. For instance, in conventional pushing beams, there are no shear waves generated in the ROE, and it is therefore impossible to find the estimates of SWS in this zone. As discussed in section 2.1, in order to reconstruct a complete image of the medium, it is necessary to generate another push pulse at different lateral positions to insonify the whole ROI with the shear waves. Alternatively, a complex pushing beam such as comb-push can be utilized. However, even in such case, a directional filter must be used to separate rightward and leftward propagating waves, which will result in two datasets being processed by a SWS estimation algorithm. Whichever case, multiple SWS maps will be found which need to be combined to build a final image. For this purpose, dedicated masks are applied to the input SWS maps, and then the masked maps are summed. Masks remove regions exhibiting estimates of low trust from the input maps. In the case, that multiple input maps contain valid data for the same regions, the final image can contain data from the selected one or a weighted sum of multiple maps [109]. An example of such an approach is shown in Fig. 2.6. A complex weighting method was proposed by Song et al. who used local energies of shear waves and correlation coefficients as weights in the weighted sum compounding [111]. To achieve smooth transitions between data from different input SWS maps, a sigmoid-based data weighting across the region boundaries can be applied [140].
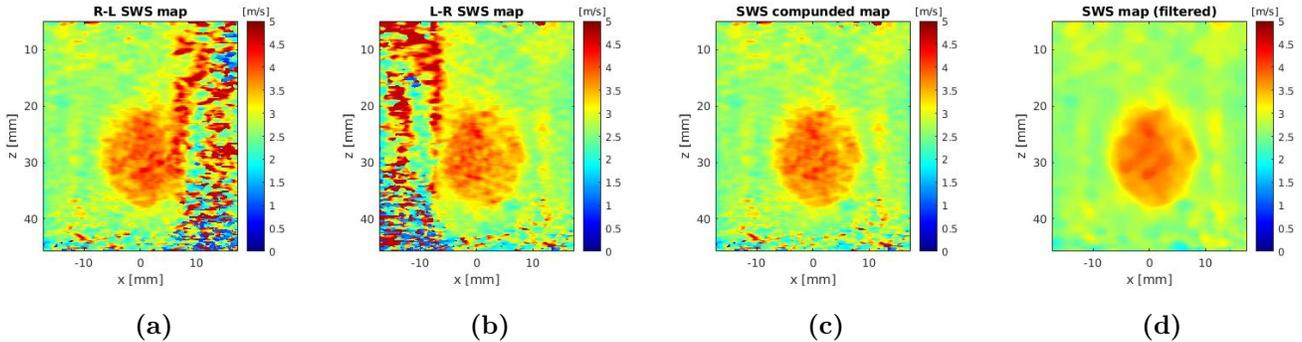
Figure 2.6: An example of SWS map compounding. (a)–(b): The SWS maps reconstructed from the shear waves that were propagating from the right to left (R-L) and left and right (L-R), respectively. Strong artifacts are visible close to the map borders. (c): The compounded SWS map formed by combining (a) and (b) and masking the regions at the margins. The regions in the middle where both input maps overlap with valid data are calculated as the weighted sum, taking correlation coefficients as the weights. (d): The final image, i.e., the result of filtering (c): using a median filter with a kernel size of 2×2 mm. Images present results obtained by the Author.

Typically some type of spatial filter is applied on the final compounded SWS map to further refine the image. In order to preserve edges, a median filter or Non-Local Means filtering can be used [133]. The final image is often presented along with a B-mode image to provide anatomy guidance. Optionally, the final SWS map can be overlaid on the B-mode image.

2.4   Safety considerations

In SWE, just as in any other ultrasound imaging procedure for clinical use, acoustic energy output can not exceed the defined thresholds to avoid thermal or mechanical bioeffects in tissue. Those limits, along with applicable definitions are imposed by the Food and Drug Administration (FDA) in the United States, while the IEC 60601-2-37:2007+AMD1:2015 standard released by The International Electrotechnical Commission (IEC) provides international requirements for the basic safety of ultrasonic medical diagnostic equipment. The discussion below will follow FDA's regulations and definitions as in [146], since they are most commonly referred in literature. Relevant acoustic measurements to characterize medical ultrasound devices are typically performed using ultrasonic hydrophones. Common procedures and recommendations for this type of characterization are described in both scientific literature (e.g. [147]), and International standards and technical specifications such as IEC 62127-1:2022 and others [148].

A)   Thermal effects

In SWE, with repetitive push pulse generation at high-voltage followed by a high frame rate tracking sequence, heating occurs either at the focal region deep in the tissue or at the transducer-skin interface. The FDA regulations limit the derated spatial-peak temporal average intensity $I_{SPTA,0.3}$ to be under $720\,\mathrm{mW/cm^2}$. It is defined as:

$$I_{SPTA,0.3} = MAX \left(\frac{1}{T_{PRF}} \cdot \int_0^{T_{PRF}} p_{0.3}{}^2(t,z)dt \right), \tag{2.9}$$

and is measured at the location of the highest value of acoustic intensity on beam axis z, T_{PRF} is the pulse repetition interval, and $p_{0.3}(t)$ is the pressure waveform versus time derated by a factor of 0.3 dB/MHz/cm, that is a conservative estimate of the overall effects of average soft-tissue absorption that not occur in hydrophone-based water measurements [4]. Additionally, to prevent thermal damage the Thermal Index (TI) is required to be lower than 6.0, which indicates that during the pulse sequence or scanning sequence, the temperature increase must be lower than 6.0 degrees. A model that allows predicting the increase of temperature as a result of energy absorption of the acoustic wave was proposed by Palmeri and Nightingale [149]:

$$\Delta T = \frac{2\alpha}{c_v} \cdot t, \tag{2.10}$$

where ΔT is the temperature rise, α is the absorption coefficient, I is the *in-situ* acoustic beam intensity, t is the heat application time, and c_v is the heat capacity per unit volume of soft tissue ($c_v = 4,200 mW \cdot s/cm^3/^oC$). While the thermal effects are negligible for single push pulses, the risk of exceeding these levels is present in real-time applications where a repetitive measurement is performed with a high duty cycle [149], [150].

B)   Mechanical effects

For the Mechanical Index (MI) which is a predictor of the occurrence of cavitation, the FDA sets the $MI_{0.3}$ limit at 1.9. It is calculated as:

$$MI = \frac{P_{r,0.3}}{\sqrt{f}}, \tag{2.11}$$

where f is the center frequency of the ultrasound wave in MHz and $P_{r,0.3}$ is the peak rarefractional pressure in MPa derated by 0.3 dB/cm/MHz to the point on the beam axis where the pulse intensity integral is maximum.

Push pulse sufficient to generate detectable displacements in soft tissue is typically achievable within those standard diagnostic limits [39], however, in multi-frame applications, reducing push generation repetition frequency may be required due to thermal reasons [150], [151]. Typically, the MI limit is more restrictive in ARF-based applications unless the scanning sequence is very long and then the TI or I_{SPTA} limit may become more restrictive [79].

2.5 Imaging quality

Images reconstructed by the 2-D SWE technique provide clinicians with intuitive maps of the tissue stiffness. Some clinical applications that benefit from the quantitative character of SWE, e.g. liver fibrosis staging, require an accurate estimation of tissue stiffness since the estimated value is compared with a pre-established quantitative scale. In such cases, an inaccurate estimation may drive a wrong diagnosis and result in improper treatment. Therefore, it is important to derive estimation accuracy. This section provides a brief overview of the parameters affecting SWE imaging quality, such as reconstructed stiffness uncertainty, bias, and image resolution.

2.5.1 Stiffness estimation uncertainty

As shown earlier in this chapter, data processing in 2-D SWE involves multiple steps. Uncertainties can arise from each algorithm's step and propagate through the data processing pipeline resulting in errors in the final reconstructed SWS maps. The further analysis assumes, that ToF algorithm is used for SWS estimation since this method was chosen for implementation in this work. As shown in section 2.3.4, in ToF method the information about local SWS is extracted from the spatio-temporal shear wave motion data. For example, in the correlation-based method, the final estimation (see Eq. 2.8) involves a calculation of both space and time parameters and each is a subject of uncertainty.

A)   Shear wave arrival time estimation uncertainty

Deffieux et al. in their seminal paper ([135]) investigated the shear wave arrival time estimation noise to find a relationship between the shear modulus estimation variance and acquisition parameters. Based on analytical derivation, numerical modeling, and experiments, they came up with a comprehensive model incorporating uncertainty coming from image acquisition and beamforming, shear wave detection, and ToF correlation-based SWS estimation. Using this model, the minimum variance σ_μ of the locally reconstructed shear modulus can be found as:

$$\sigma_\mu^2(x,z) \approx \frac{\mu_0(x,z)^3}{\Delta d^2} \cdot \frac{1}{2\pi^2 F_{SW}^2 BW_{SW} \cdot T}$$
$$\cdot \left[\left(1 + \frac{1}{SNR_{v_z}\left(x - \frac{\Delta d}{2}, z\right)} \right) \cdot \left(1 + \frac{1}{SNR_{v_z}\left(x + \frac{\Delta d}{2}, z\right)} \right) - 1 \right], \tag{2.12}$$

where μ_0 is the true local shear modulus, Δd is the lateral SWS estimation kernel, T is total data acquisition time, F_{SW} and BW_{SW} are shear wave mean frequency and shear

wave bandwidth, respectively. The SNR of the shear wave motion data, SNR_{v_z}, is non-uniform within the FOV and can be calculated by shear wave axial velocity energy divided by shear wave axial velocity variance σ_{v_z} [135]:

$$SNR_{v_z}(x, z) = \frac{\int_T v_z(x, z, t)^2 \, dt}{\sigma_{v_z}(x, z)^2},$$
(2.13)

and

$$\sigma_{v_z}(x, z) = \frac{c^2 \cdot FPS^2}{64 f^2} \cdot \frac{1}{\pi^2 r^2} \left(1 - r^2 + \frac{2}{SNR_{Uf}(x, z)} \right),$$
(2.14)

where FPS is frame rate, $v_z(x, z, t)$ shear wave axial velocity profile along slow-time t, f is tracking pulses frequency, c is speed of sound, $SNR_{Uf}(x, z)$ is SNR of ultrasound frame at a given pixel, and r is the correlation coefficient between successive frames in a noiseless environment resulting from the stochastic nature of ultrasonic speckle. Its derivation can be found in [135]. In practice, rather than applying Eq. (2.14), $\sigma_{v_z}(x, z)$ can be obtained experimentally.

Analysis of the above formulas reveals that:

- Higher frequency and wideband transient shear waves yield more precise estimation. Generation of shear waves of the same energy but with wider bandwidth would require a shorter pulse with higher intensity. Moreover, shear waves of higher frequencies would suffer excessive attenuation, which can result in a reduced propagation path.

- The estimation variance is related to the true stiffness of the medium, which is unknown. The higher the SWS is, the higher is estimation variance.

- Estimation precision can be improved by increasing the lateral SWS estimation kernel size Δd but at the cost of resolution loss.

- The SNR of shear wave motion data is higher for higher tracking pulse frequency, however, higher frequency attenuation may also limit the SNR of ultrasound frames at larger depths.

- With lower frame rates the amplitude of shear wave axial displacements is higher between frames, and as a result, shear wave detection variance can be improved in the presence of noise. However, it is done at the cost of reduced temporal resolution and lower correlation values [135].

Although the model of Deffieux described above assumed usage of correlation of adjacent shear wave axial velocity profiles, other works verified, that other algorithmic approaches to obtain shear wave time arrival suffer similar levels of precision [139].

B)   Speckle bias

Although named as bias, speckle bias is a source of shear wave arrival time estimation uncertainty, that is not included in the model of Deffieux mentioned above. It was described and analyzed by McAleavey et al. [152]. Speckles provide a texture for motion tracking throughout the tissue. McAleavey et al. observed, that when tissue motion is tracked, the system is more sensitive to the areas of bright speckles. As a result, the shear wave arrival times at particular tracking locations tend to bias toward the bright

speckles. In other words, the speckle bias manifests as an apparent variable spatial offset in the location of the tracking positions, which is correlated with the local stationary speckle pattern. Since the SWS is estimated as a ToF over the distance between tracking locations, the uncertainty in the tracking positions resulting from speckle bias affects the SWS estimation. The magnitude of this bias is determined by the size of the Point Spread Function (PSF) of the system, as speckle size is correlated with PSF size [152]. Speckle bias leads to lateral distance errors up to half of the PSF lateral size [107]. It can be also simply derived, that the magnitude increases along with the stiffness of the medium [152].

One method to suppress the speckle noise is to use a larger lateral kernel for performing SWS estimation. However, such an approach would reduce the spatial resolution of the reconstructed stiffness maps. Thus, there is an inherent trade-off between lateral resolution and suppression of the speckle bias, which may limit the effective resolution to a few times larger than the speckle size [153]. In addition, averaging data from multiple acquisitions may suppress arrival time jitter caused by electronic noise, but cannot diminish the stationary speckle bias, since it is a correlated noise.

These properties of the speckle bias motivated the development of Single-Tracking Location (STL) shear wave elastography [66], [154]. By taking the difference of arrival time estimates of shear waves generated by two push beams separated by a known lateral distance and tracked at a single location, the STL method is almost immune to speckle noise. Hollender et al. compared common multiple-tracking locations (MTL) and STL methods in terms of contrast and resolution, showing that the STL method can significantly outperform MTL in terms of image quality, creating less noisy and higher-resolution images [153]. However, it requires many acquisitions to reconstruct a single image, which significantly increases acoustic exposure and reduces frame rate, exposing this technique to potential motion artifacts and making it less suitable for real-time applications.

C) *Phase aberration*

The assumption of a constant 1540 m/s speed of sound (SoS) in the tissue is used in ultrasound systems to compute delay profiles to be applied for electronic steering and focusing with array transducers. *In-vivo* soft tissues are inhomogeneous, with intrinsic SoS variations. Wavefront propagating through such tissue can exhibit phase aberration. In the presence of phase aberration, the propagation delays computed for beam focusing are no longer valid, which can result in main lobe energy spreading, higher sidelobes amplitudes, and distortion of PSF shape [107]. All those effects degrade image quality. In the case of SWE, steering errors coming from phase aberration can lead to lateral shifts of the tracking beams. Errors of lateral positions coming from these shifts result in SWS estimation errors. Deng et al. performed 2-D acoustic propagation simulations to demonstrate the effect of phase aberration [107]. The simulations were designed to image a point target under the 3 cm width abdominal wall consisting of mostly fat of SoS of 1479 m/s. The results have shown, that using 2 MHz transmit frequency at F/2 focal configuration to image a point target located at 6 cm depth, its lateral position errors ranged up to 0.33 mm, depending on the abdominal wall model used. Substantial SWS estimation errors resulting from phase aberration were also reported by Chintada et al. [155].

The phase aberration effect can be partially suppressed by using harmonic imaging methods for shear wave tracking [119], [120]. In [155], the authors proposed using local

SoS adaptive beamforming to compensate for phase aberration.

2.5.2 Bias

Similarly to uncertainty, biases in the SWE technique can arise from both spatial and temporal sources. These bias effects were investigated in works by several research groups.

In a systematic study, Zhao et al. demonstrated that SWS estimates can depend on the transducer, depth, and lateral tracking position [156]. They have shown experimentally, that an undesired push beam intensity field could be a possible cause of bias; wide beam widths in the elevation direction can generate significant out-of-plane shear waves that bias the SWS estimates high. It was reported that this effect is stronger in close proximity to the push beam axis in the lateral direction. It also increases with distance from the beam focus in the axial direction. In the same work, they also speculate, that diffraction could be another cause of the bias. For a pushing beam, diffraction produces acoustic intensity before and beyond the depth of focus, which generates additional shear waves in non-focal regions and interferes with shear waves from the focal point, thereby biasing the results high. This effect was found stronger for wider apertures. All aforementioned potential sources of bias are associated with ToF-based SWS estimators relying on *a priori* information about the shear wave propagation direction from a focused beam; the shear waves are assumed to propagate only in a direction perpendicular to the pushing beam axis. Song et al. described the 2-D SWS estimation scheme to partially solve this issue [111]. In this approach, a curvilinear probe was used and SWS estimates were obtained along both the lateral and axial dimensions, which allowed to obtain SWS of shear waves traveling at the oblique angles.

Biased results could be also caused by system-dependent parameters, such as PRF error and transducer parameter mismatches between the real values and those used in the software. Both factors are linearly related to the corresponding SWS estimation bias [107]. For example, a 5% pitch error would result in a 5% SWS estimate error. The same applies to PRF errors. However, both factors can be neglected by a careful system calibration [123].

Another source of bias can be present in the case of imaging the inclusions of diameters comparable to or smaller than the shear wavelength. In such a case, transverse wave propagation is distorted by local heterogeneities and boundaries that can bias shear and the local medium homogeneity assumption (see Chapter 1) is no longer valid. Consequently, accurate inclusion stiffness measurement is not possible; Shear wave cannot propagate at the speed reflecting real inclusion stiffness unless the whole shear wave enters the inclusion region, and thus the local homogeneity assumption holds [133]. Song performed numerical simulations and experimental tests of elasticity imaging of inclusion embedded in softer background material [133]. He observed, that as the inclusion size gets smaller, the apparent stiffness of the inclusion gets softer (negative bias). The same observations were made by Racedo et al. [140] and Cacko et al. [80], who reported the levels of SWS bias due to this effect ranging from -40% to +5%, depending on inclusion nominal stiffness and diameter. In order to minimize this effect, a sharp shear wave can be generated, which is, however, very challenging due to the viscoelastic nature of tissue that tends to substantially broaden the shear wave [133].

2.5.3 Resolution

The resulting stiffness maps resolution in terms of image pixel size is determined by the ultrasound imaging resolution and is not related to SWE resolution. In the same way, as ultrasound imaging spatial resolution is largely determined by the length of the transmitted pulse, the spatial resolution of SWE is largely determined by the shear wave wavelength. The SWE resolution is often expressed as the minimum inclusion size that can be detected. This approach is considered a relaxed condition in defining spatial resolution, in which the SWS of the object does not have to be accurately determined [133]. For example, inclusion can be still detected when the shear wave wavelength is larger than the inclusion size, nonetheless, the obtained inclusion stiffness may be biased as discussed in the previous section. This way shear wavelength can be used to roughly estimate SWE spatial resolution. Still, it should be noted, that shear wavelength depends on both excitation parameters and medium characteristics [41]. Typically, inclusions of minimum diameters 4 to 6 mm can be detected, depending on inclusion and background stiffness contrast and reconstruction parameters [80], [112], [133], [140].

A more analytical approach to resolution metric was used by Rouze et al., who fitted the reconstructed SWS profiles across the inclusion using a sigmoid function [139]. The resolution was then assessed using the slope of the fitted function. However, in some cases, this approach failed in the presence of shear wave reflections at the inclusion boundaries and suffered significant uncertainty. Even though, using this method, Rouze et al. performed numerical and experimental studies which have shown, that the key factors affecting the SWE resolution are shear wave excitation beamwidth and lateral SWS estimation kernel size. Regarding kernel size selection there is a tradeoff between SWS estimation precision and spatial resolution, i.e. increasing the kernel size improves estimation precision since the SWS is averaged at a greater distance, but is done at the cost of reduced resolution.

Chapter 3

Shear wave elastography implementation on a portable research system

As discussed in previous chapters, the 2-D SWE is a sophisticated technique that poses several technical challenges that need to be solved for successful implementation. This chapter describes the efforts targeted to overcome these challenges in the practical implementation of this imaging mode using a portable ultrasound system. The description of this process is divided in this chapter into three parts: hardware development, image reconstruction algorithm development, and implementation of real-time operation.

3.1   Hardware development

The us4R-lite research system (us4us Ltd., Warsaw, Poland) shown in Fig. 3.1 was used as a hardware platform for the implementation of the 2-D SWE mode. Firstly, this system architecture is briefly discussed. To achieve the goals of this thesis, the transmit capabilities of this system required to be enhanced by the development of a dedicated transmit extension module allowing push pulse generation for SWE. A description of the transmit extension module hardware design and FPGA firmware architecture constitutes the remaining part of this section.

3.1.1   us4R-lite research system overview

The us4R-lite system architecture is consistent with a typical ultrasound software-based open platform organization — as defined by Boni et al. in [9] — to effectively facilitate the practical evaluation of new ultrasound imaging methods. In the case of the us4R-lite system, it is achieved by the operational programmability of transmit (TX) and receive (RX) operations to support both conventional and alternative acquisition sequences. The us4R-lite scanner provides access to raw RF channel data and the data processing can be flexibly programmed for a specific application. All these features combined make this device a suitable platform for the deployment of ultrasound imaging algorithms that require custom TX/RX sequences and custom data processing. The 2-D SWE is an example of such a technique.

Figure 3.1: us4R-lite portable ultrasound research platform with a commercial probe connected.

A) System architecture

The top-level architecture of the us4R-lite research platform is depicted in Fig. 3.2. The system is built around the two us4OEM ultrasound transceiver modules, each containing both TX and RX front-end electronics. The single module has individually controlled 128 TX and 32 RX channels, giving 256 TX and 64 RX channels in total for a complete system. Transmit part is based on 3-level transmit pulsers integrated circuits (ICs) that drive individual probe elements with high-voltage (HV) bipolar square waves. The receive path consists of an amplifying stage (low-noise amplifier (LNA) and time-gain control amplifier, (TGC)) followed by a 14-bit analog-to-digital converter (ADC) that samples the analog signal at 65 MSPS. A complete analog receive circuit along with the ADC is integrated within a multichannel analog front-end (AFE). Transmit and receive paths are connected in the transmit/receive (T/R) switch, which connects the transducer elements to either transmitter or receiver subsystems by multiplexing. Both TX and RX parts are controlled by the Field Programmable Gate Array (FPGA) configured with custom firmware. The FPGA is also responsible for the real-time execution of a pre-defined TX/RX acquisition sequence which is programmed by the software prior to the acquisition. For TX operations, the FPGA generates the transmit patterns to be amplified by TX pulsers and sent through each array element. Regarding RX handling, the FPGA captures output data streams from ADCs, which are then transferred temporarily to the off-chip buffer RAM. All mentioned components are mounted on the multilayer printed circuit board (PCB) of the us4OEM modules. Other components that build the system include: clock generation and synchronization circuits, a power supply that generates all the required voltage levels for other subsystems, and an HV power supply that generates bipolar high-voltage supplies required for the transducer excitation during the TX phase. The power supplies and clocks are shared by both us4OEM modules. All the subsystems are connected by digital and analog backplanes. The probe adapter is an exchangeable interface to the ultrasound probe. Typically each probe requires a dedicated probe adapter to provide proper connections to the system.
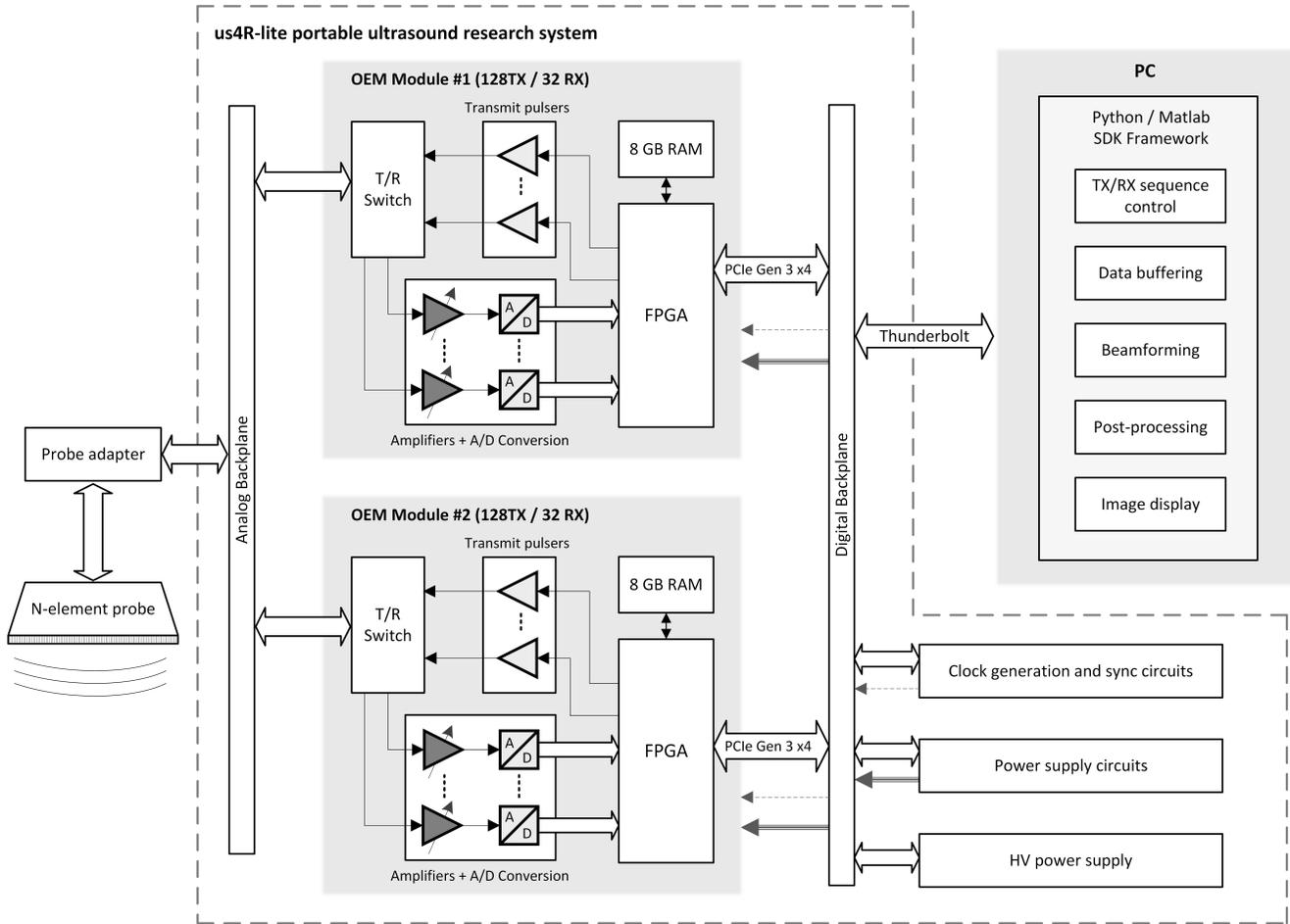
Figure 3.2: The us4R-lite portable ultrasound research platform simplified system architecture. Some connections are hidden for clarity. Abbreviations used: T/R - transmit/receive, A/D - analog-to-digital, FPGA - Field Programmable Gate Array, SDK - Software Development Kit, HV - high voltage.

B) Data transfer and processing

Since there is no hardware-based data processing, all the acquired RF data are fed to the PC-based computing back-end. Lack of beamforming results in a large volume of data being transferred to the PC, especially in real-time applications [157]. For this reason, the us4R-lite system is connected to the PC using a high-speed PCIe Gen3 x8 or Thunderbolt-3 interface providing data bandwidth of 6 GB/s or 2.4 GB/s, respectively. The computing station working with the system can be any PC supporting PCIe or Thunderbolt interface. Due to the large amount of data to be processed, the workstation should also feature sufficient memory and processing resources. To achieve sufficient performance, both central processing units (CPUs) and graphics processing units (GPUs) are often utilized to benefit from massive parallel processing [9]. Since the introduction of high-level GPU application programming interfaces (APIs) such as Nvidia CUDA (Nvidia, Santa Clara, USA), a realization of computationally intensive ultrasound imaging algorithms in various fields has been reported by many authors, e.g.: beamforming [158]–[161], color Doppler imaging [162], [163], utilization of deep learning in ultrasound [164], and also displacement tracking in transient elastography [165].

Although GPUs can handle all the data processing from raw channel data to the final image, in hybrid systems containing CPUs, FPGAs, and GPUs, the strengths of each of these devices can be combined for more efficient computation [7], [9], [166], [167]. Design partitioning between the hardware and software is considered to be an important engineering topic in modern systems executing complex imaging algorithms in real-time [9], [98].

In this work, the MSI GS66 mobile workstation (Micro-Star Int'l Co. Ltd., Taipei, Taiwan) was used as a host connected to the us4R-lite system, equipped with Intel Core i9-10980HK CPU with 18 cores, Nvidia RTX3060 GPU, and 32 GB of RAM memory.

C) Author's contribution to the device

The us4R-lite research platform is a complex device and several authors contributed to its final form by bringing efforts in areas of many specialties, such as electronic circuits design, PCB design, FPGA digital design, MCU firmware development, mechanical design, software development, validation, and testing. As a portable device that is substantially smaller than the competition, its construction featured many engineering challenges that needed to be solved, in particular, related to a high density of elements, electromagnetic compatibility (EMC) issues, and extensive heat dissipation at limited space. The author of this thesis contributed to the FPGA design part of this system, by developing the whole TX-related digital logic. He also developed a central sequencer IP (Intellectual Property) Core, which is responsible for real-time execution control of the pre-defined TX/RX acquisition sequence.

3.1.2 TXPB-256 transmit extension module

Despite promising results of us4R-lite system evaluation towards push pulses generation in previous works [168], multiple TX pulsers chips were damaged during preliminary SWE experiments held for the purpose of this dissertation. Possible reasons were HV overshoots at TX outputs or overheating due to long transmit patterns. Therefore, in order to increase the robustness of the us4R-lite system transmit capabilities, it was decided to design a new, dedicated 256-channel transmit subsystem — the Transmit Push Beamformer Board (TXPB-256). It has been implemented (Fig. 3.3) in the form of an extension module that comes in place of the Analog Backplane (see Fig. 3.2), effectively superseding the transmitters and T/R switches on the us4OEM boards. The block diagram of the architecture of the TXPB-256 board is shown in Fig. 3.4. This module was the key element that allowed the implementation of the 2-D SWE technique. Contrary to the us4R-lite platform, which is a general-purpose system designed by a group of people, the TXPB-256 module was designed solely by the Author specifically to make the goals of this thesis achievable. Therefore it is described in detail below.

A) Detailed hardware architecture

The TXPB-256 module utilizes the new generation 16-channel STHV1600 transmit pulsers ICs [169] (STMicroelectronics, Geneva, Switzerland) as the transmit excitation device. The module contains 16 of those chips, building a 256-channel transmit solution. The documentation of the STHV1600 pulsers is confidential and is shared under the non-disclosure agreement (NDA). Therefore, only a general overview of this device is provided below.

<div align="center">(a)   (b)</div>
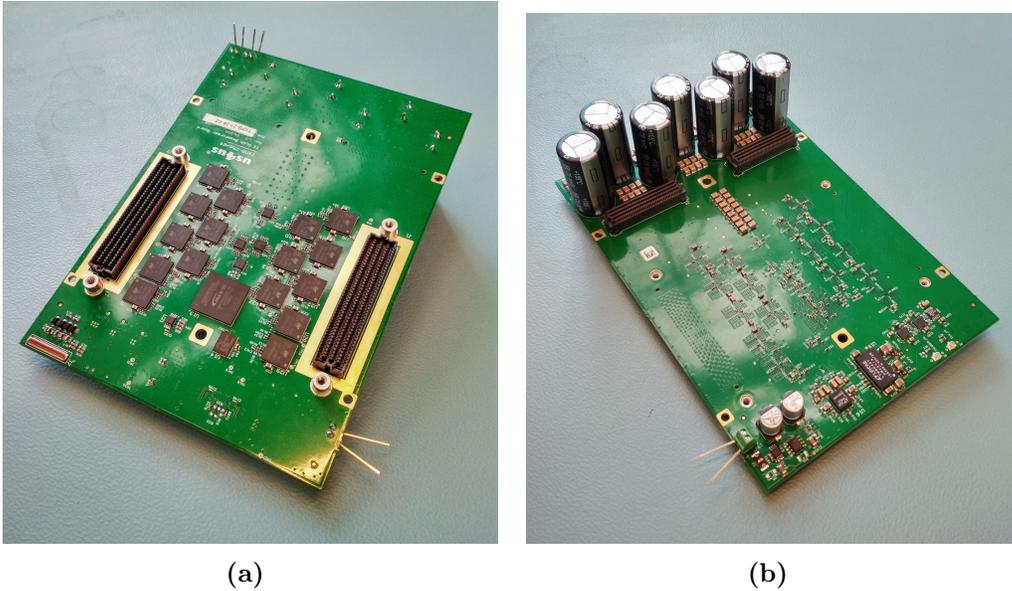
Figure 3.3: The TXPB-256 transmit extension module. (a) Top side. (b) Bottom side.

Each monolithic pulser integrates analog and digital parts. The analog part implements an HV output drive circuit that hosts two independent transistor half bridges capable of up to ±2A peak output current in 3 or 5 levels configuration. If putting these bridges in parallel, outputs can provide 3-level profiles up to ±4A current capability on all 16 outputs. Each analog channel also integrates an active clamp circuit and a T/R switch guaranteeing high isolation between the HV transmitter and the low-voltage receiver during a transmit phase. During a receive phase, the T/R switch is turned on and provides a low resistance connection between the transducer and the receiver input.

The digital part of the TX pulsers embeds a digital transmit beamformer, RAM, and a set of registers to configure excitation settings. Each TX channel can be written with a set of arbitrary 3- or 5-level waveforms. The waveforms can differ between channels. Each channel is equipped with a set of registers, which define a transmit delay, transmit enable and a waveform to be picked for the next TX event. The waveforms RAM and registers can be accessed by an external controller using a dedicated Quad Serial Peripheral Interface (QSPI). An excitation begins upon the rising edge of the externally supplied trigger signal and the transmit generation is controlled by the embedded digital beamformer, according to the pre-defined settings. Except for waveforms, the pulsers can store only a single TX definition, therefore, there is an onboard FPGA device, which reloads the definitions during runtime to execute complex TX/RX schemes. The FPGA operation is described in detail in the next section.

Besides TX pulsers and FPGA, the TXPB-256 module integrates auxiliary components and circuits. Three external power supply rails are required for proper operation: symmetric ±HV voltages and the main 19 V to power. Power supply voltages required by each individual component are provided by the onboard power supply circuit. Using 19 V input voltage, it generates 3.3 V, -3.3 V, 1.8 V, and 1.2 V routed throughout the board. The power supply circuit is based on LTM4622 [170] (Analog Devices, Wilmington, USA), dual step-down DC/DC converter, Analog Devices LTM8049 [171], dual SEPIC/Inverting DC/DC converter and linear regulators: ADP1755 [172] and
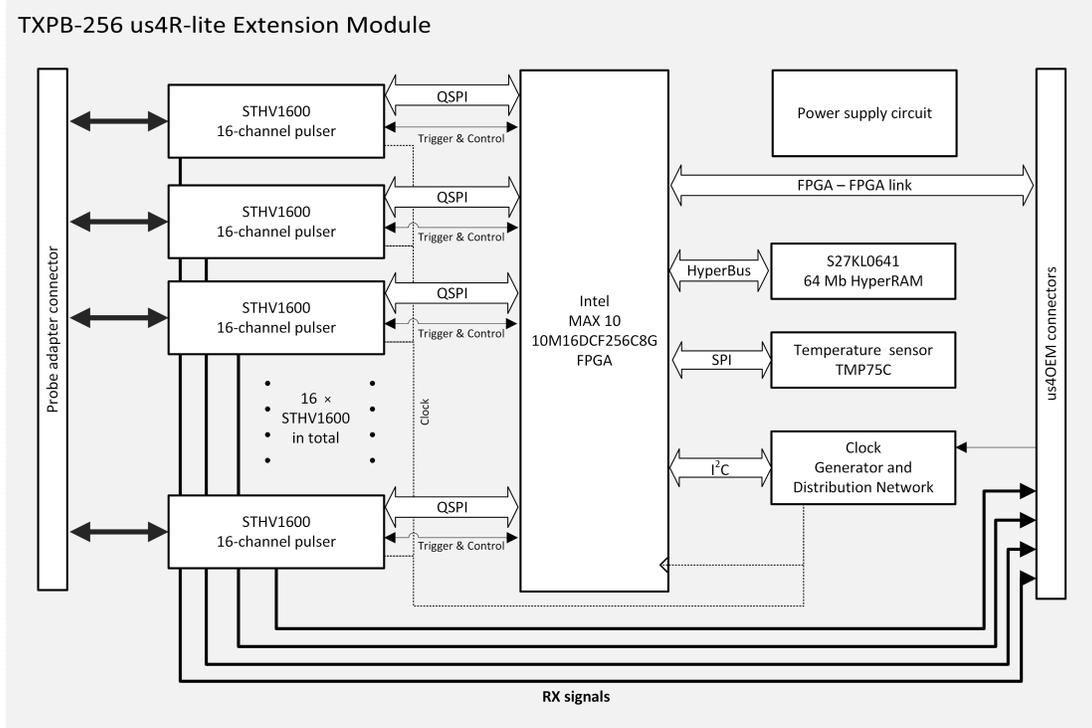
Figure 3.4: Transmit Push Beamformer (TXPB-256) extension module architecture.

LT3015 [173] by Analog Devices. In addition, there are massive electrolytic capacitors (7 x 1 mF) that decouple HV rails to avoid excessive voltage drops on the HV supplies due to high current consumption during push pulse generation. To ensure synchronous operation of digital components, there is a Si5338 clock generator [174] (Silicon Labs, Austin, USA) that uses a 65 MHz input clock to generate a 130 MHz clock for FPGA and 195 MHz clock for TX pulsers. Clocks are distributed by the buffered network built by Si5330 [175] and Si53342 [176] low-jitter clock buffers by Silicon Labs. To increase memory space for transmit-related definitions storage, the module integrates S27KL0641 64 Mb HyperRAM memory [177] (Cypress Semiconductor, San Jose, CA, USA). There is temperature sensing of the board provided by the TMP75C digital temperature sensor [178] (Texas Instruments, Dallas, TX, USA). All aforementioned components are mounted on the carefully designed 14-layer PCB.

B) Transmit capabilities

The transmit capabilities of the extension module are summarized in Table 3.1. The high configurability of transmit parameters provides flexibility in designing transmit patterns. For a given transmit event, each channel can be turned on or off and can have an individually applied excitation delay. In particular, each channel can output a different waveform. However, although an arbitrary pattern generation is possible, the pulse shaping is constrained by only 3 possible levels — +HV, -HV, and clamp to the ground — and by a timing precision of 5 ns. However, the flexibility of the signal design can be increased by the use of pulse width modulation (PWM). PWM can be used effectively to produce complex analog waveforms with bandpass filtering by the transducer itself [179]. This approach allows in the developed system to perform

62

Table 3.1: usR-lite research platform electrical transmit capabilities with (TXPB-256) extension module.

Feature	Value
TX channels	256
HV transmit voltage	Configurable, up to 180 Vpp
Output current per channel	$\pm$2A or $\pm$4A
Transmit frequency range	0.1–50 MHz[1]
Transmit aperture	Arbitrary
Transmit pulse capabilities	Arbitrary 3-level square wave pattern for each channel
Transmit apodization	Yes[2]
Transmit delays	0-327 μs
Timing resolution	5 ns
Theoretical maximum PRF	100 kHz

[1] Maximum frequency is load dependent.

[2] Implemented using pulse width modulation. Details are in the text below.

aperture apodization by setting individual transmit waveform amplitude levels over the active aperture elements. The amplitude of the fundamental frequency component of the wave, A, is dependent on PWM factor M in a non-linear relationship [180]:

$$A = A_0 \cdot \sin\left(M \cdot \frac{\pi}{2}\right), \tag{3.1}$$

where the A_0 is the amplitude of the signal at full pulse width ($M = 1$). It should be noted, that due to limited timing precision, the number of possible amplitude levels is also limited — the higher the frequency of the signal, the lower the number of possible amplitude levels. For example, with a signal frequency of 5 MHz and timing precision of 5 ns, there are only 11 amplitude levels available.

3.1.3 TXPB-256 module firmware design

The main part of the TXPB-256 module is the transmit stage built with 16 STHV1600 TX pulsers. As pointed out in the previous section, these devices contain single transmit definition registers. Since transmit definitions may differ for each transmission, they need to be reloaded using the QSPI interface between TX events to execute a complex transmit strategy. To ensure that the transmit PRF is not limited, it is essential to load those definitions as quickly as possible. For this purpose, a dedicated Intel MAX 10 FPGA device (10M16DCF256C8G) was placed on the TXPB-256 module. This device acts as a high-speed control unit, which reconfigures the TX pulsers dynamically during runtime, depending on a pre-defined transmit plan for desired imaging scenario. This process is described later in this section. Other tasks that the FPGA is responsible for include:

- Providing read/write access to internal configuration registers and memory of TX pulsers using QSPI interfaces for configuration of constant parameters or reading the status registers.

- Providing read/write access to onboard HyperRAM memory to store transmit-related definitions to be used during execution of imaging scenario.

- Providing read/write access to the onboard temperature sensor for configuration and read-out of temperature measurement results.

- Providing read/write access to on-board clock generation circuit to control generated clock frequencies or allow clock gating for power savings. The FPGA also performs clock generator initialization upon system power-up.

- Capturing and handling interrupt signals coming from TX pulsers that indicate transmit failure, wrong configuration, or other issues.

A)   Firmware architecture

The general TXPB-256 FPGA firmware architecture with major IP components and control flow is pictured in Fig. 3.5. The FPGA architecture contains several functional IP blocks, each intended to perform different tasks. All the blocks build the system connected by the internal Intel Avalon bus interconnect, which solves the communication within the system providing address decoding and data bus multiplexing. All the IPs use standard memory-mapped bus interfaces (Avalon-MM) for easy integration. For proper operation, the FPGA requires an external controller, which performs the FPGA configuration via the custom QSPI interface for the desired transmit plan. Table 3.2 gives an overview of the functionality of each module implemented within the FPGA.
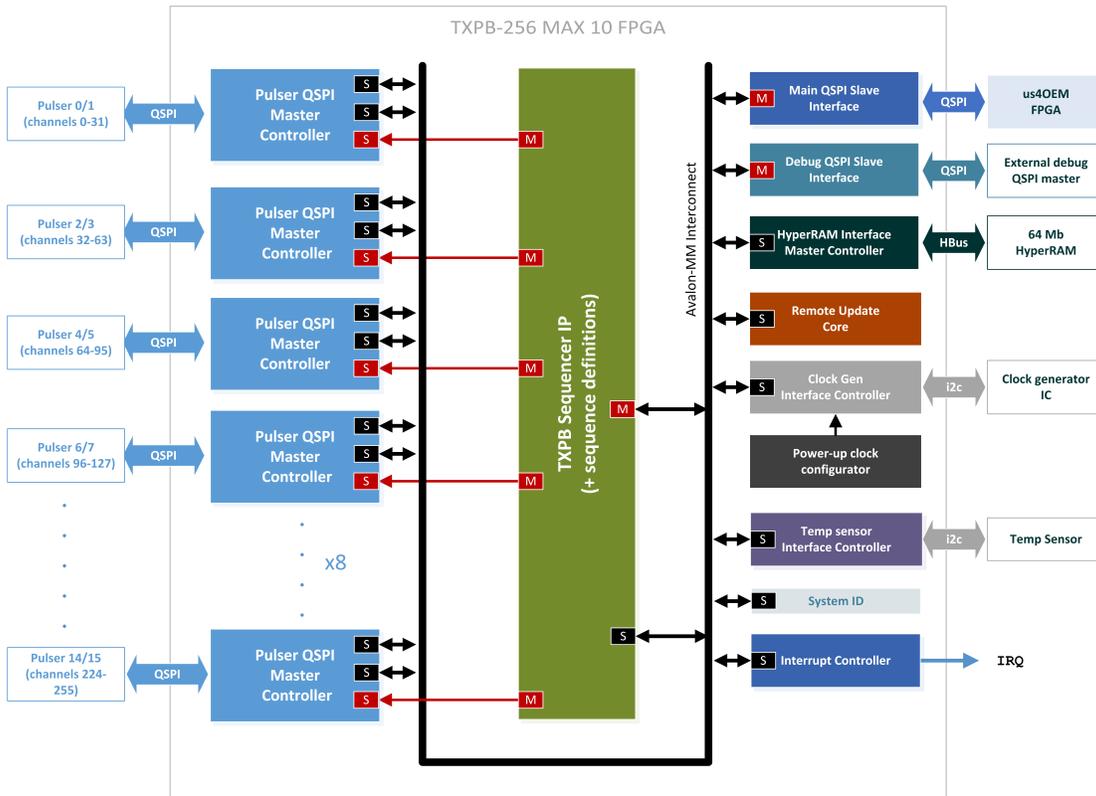
Figure 3.5: Block diagram of TXPB-256 onboard Intel MAX 10 FPGA firmware. All IP blocks are shown with interfaces: "M" stands for Avalon-MM master, whereas "S" stands for Avalon-MM slave. A brief description of all the blocks is provided in Table 3.2.

Table 3.2: TXPB-256 FPGA internal functional IP blocks.

IP Block	Description
(Main / Debug) QSPI Slave Interface	This block implements the custom protocol QSPI interface to access FPGA internal register space. It translates custom QSPI transfers into Avalon bus transactions. This block is used as an entry point for the external controller — using this interface, the external controller has access to all other functional blocks within the system. There are two identical blocks in the system — one for communication with FPGA on the us4OEM module and the second one to handle communication for debug purposes.
HyperRAM Memory Interface Controller	This block is an interface to an external HyperRAM dynamic RAM 64 Mb chip, optimized in terms of area and performance. It performs bridging between Avalon and HyperBus interface protocols.
TXPB Sequencer IP	TXPB Sequencer IP is the key component of the system. It contains the control logic to execute the TX/RX sequence pre-defined by the user. The Core provides Avalon-MM slave interface configuration memory, control, and status registers. It also features the Avalon-MM master interface to download new excitation definitions from external memory-mapped HyperRAM memory. This is done automatically by internal logic. A detailed operation of the core is provided in the text.
Clock Gen Interface Controller	This core is an interface between the Avalon bus and the I2C interface to control and configure the onboard Si5338 clock generator.
Temp sensor Interface Controller	This core is an interface between the Avalon bus and the I2C interface to control and configure the onboard TMP75C temperature sensor device.
Power-up Clock Configurator	This is the Nios2 CPU-based subsystem which is responsible for the configuration of the onboard Si5338 clock generator during power-up using the predefined configuration bitstream. It uses the Clock Gen Interface Controller to write and read internal registers of the Si5338 chip. As the main clock source, it uses an internal FPGA oscillator. The CPU code is written in the read-only RAM, which contents are defined during firmware design synthesis.
Interrupt Controller	The interrupt controller captures, masks, and handles the interrupt requests from the interrupt sources within the entire system. It produces a single signal that drives an external system interrupt output port.
TX Pulser QSPI Master Controller	This core is an interface to the STHV1600 pulser using its custom QSPI interface. It provides both simple read/write transfer mapping between Avalon and QSPI and more sophisticated configuration schemes of downstream chips. In particular, it implements dedicated burst transfers to configure all the transmit-related definitions in minimum time. Each interface is connected to two STHV1600 pulsers.
System ID	System ID is a simple read-only register that stores the revision number of the FPGA firmware. It can be used to identify the firmware version during runtime to ensure hardware-software compatibility.
Bus Interconnect	The bus interconnect is an infrastructure that joins all the modules in a complete system. It performs address decoding, pipelined access scheduling, simultaneous access arbitration, and data bus multiplexing and features multi-master access handling for optimized data transfers within the system.

B)   Main sequencer operation

As mentioned above, the dynamic reconfiguration of the TX pulsers is the key function of the FPGA. The most important system component that fulfills this task is the TXPB Sequencer IP, which simplified structure is presented in Fig. 3.6. It has three internal memories: one FIFO (First-In-First-Out) memory for firing buffering, one for TX sequence definitions storage (Sequence Table RAM), and one for control registers. The term *firing* introduced here for simplification is meant as a complete set of definitions related to a single TX event. In this case, a firing consists of:

- Transmit delay profile — a set of excitation delays for each transmit channel within the active aperture,

- Transmit aperture — a set of transmit enable settings for each channel,

- Receive aperture — a set of receive enable settings for each channel,

- Wave configuration — picks a pre-defined transmit waveform to be generated for each transmit channel.

The definitions of firings components are stored in HyperRAM memory. The Sequence Table contains entries, that define a firing for each TX event — it stores information on which transmit delay profile, which transmit and receive apertures, and which wave configuration to pick for a given TX event out of all possible definitions stored in memory. In addition, each Sequence Table entry contains fields that define an entry to be executed next and the number of repetitions of the current entry.

Figure 3.6: Sequencer IP internal architecture. The operation of the core is described in the text.

The operation of the TXPB-256 Sequencer IP core is as follows. Upon capturing the trigger event, the main sequencer state machine reads the Sequence Table RAM and then proceeds to download the firing definition from an external HyperRAM memory to the internal FIFO buffer. Then, the TX Pulser Configurator block detects data available

in FIFO and proceeds to load the memory contents to the TX QSPI Controllers for further processing. The TXPB Sequencer IP also generates control signals, like triggers for TX pulsers and TX QSPI Controllers. After the current TX/RX event is done, the core triggers TX QSPI Controllers to push new firing data to the pulsers to be ready for another TX event. This pipelined architecture of the logic allows operation with very high PRF.

C) *TX/RX configurability*

In summary, the TXPB-256 transmit configurability can be described as follows:

- Configurable transmit pulse design:

  - many arbitrary transmit waveforms storage; the number of definitions depends on waveforms' complexity (up to 256 unique waveforms),
  - configurable transmit waveform selection independently for each channel — up to 1024 wave configurations (unique definitions of channel/waveform pairings),

- Configurable transmit/receive aperture:

  - Up to 4096 unique transmit and receive aperture definitions,
  - Each individual transmit channel can be enabled or disabled for transmission/receive,

- Configurable transmit delay profile:

  - Up to 1024 unique transmit delay profile definitions,
  - Transmit delay set individually for each transmit channel,
  - 16-bit delay definition, setting resolution 5 ns.

- Configurable TX/RX sequence:

  - up to 1024 entries in the sequence table configurable by the user,
  - fully configurable entries execution order,
  - each sequence table entry picks:
    * a single (out of all possible definitions) wave configuration,
    * a single transmit aperture to be applied,
    * a single receive and aperture to be applied,

D) *Design implementation and verification*

Although FPGA firmware design implementation and verification tasks consumed a huge amount of effort and time, they were found too technical to be described in detail in this dissertation. Therefore, only a brief note on methodology is written below.

Most of the firmware components were custom-designed IPs written in VHDL (Very High-Speed Hardware Description Language). Thus, most of the logic is easily transferable between FPGA devices of different families or even from different manufacturers. Design implementation involved standard steps of logic synthesis, mapping, placement,

and routing that were all performed using Quartus Standard Edition 19.1 software (Intel, Santa Clara, CA, USA).

Since FPGA operation is hard to debug, a lot of effort was put into design verification to ensure that logic operation meets the specification. The design verification was performed in two steps. The first one involved functional simulation using Modelsim software (Siemens, Munich, Germany). For this purpose, a testbench was designed in VHDL for each IP component that included stimulus and functional models of external devices. After that, the whole design was also simulated. Various test scenarios were simulated to achieve good simulation coverage. In a second step, a SignalTap (part of Quartus software) integrated logic analyzer was inferred into logic to track the internal operation of the working FPGA. This step was performed after FPGA firmware deployment in the target device and required the development of test scripts that configured the FPGA into operation scenarios being the subject of tests.

E) Resource utilization and performance

The resource utilization of the FPGA firmware design is shown in Table 3.3. The design fits into the chosen target device. Regarding performance, the design met the timing closure for the 130 MHz input clock at worst-case operating conditions.

Table 3.3: The TXPB-256 FPGA resource utilization. Acronyms used: PLL - Phase-Locked Loop, I/O - input/output.

Resource type	Available	Usage	%
Logic elements	15,840	11,258	71%
Flip-flops	16,679	7,396	44%
18x18 multipliers	45	0	0%
M9K (9 Kb) memory blocks	64	46	75%
PLLs	4	1	25%
Dedicated clock networks	20	9	45%
I/O ports	178	143	80%

3.2   Acquisition configuration software

Configuration of the us4R-lite system with the TXPB-256 transmit extension module TX/RX sequences for custom data acquisitions involves applying many settings in a number of subsystems and individual devices. Therefore, for convenience, this process was automated by a set of software scripts. The general acquisition configuration software organization is presented in Fig. 3.7. It consists of four components and each is described below.

- **TXPB-256 HAL Driver**

  The TXPB-256 hardware abstraction layer (HAL) driver provides a set of scripts in Python that allow communication and configuration of the TXPB-256 transmit extension module. User can use high-level constructs to describe desired TX sequences. HAL driver automates tasks such as the calculation of transmit delay profiles, transmit waveforms design, apertures configuration, and configuration of low-level hardware registers that control the execution of custom TX sequences.
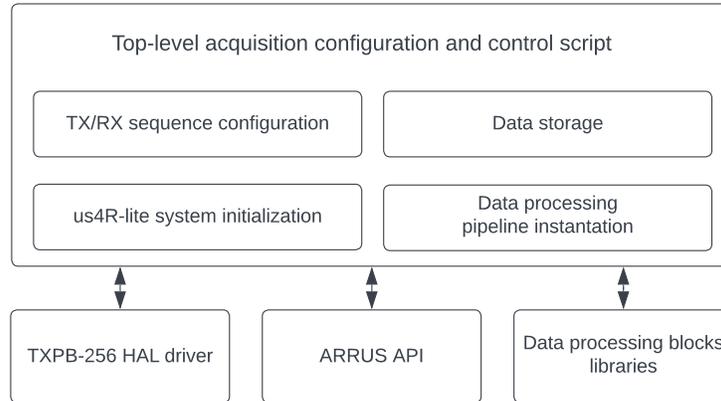
Figure 3.7: Top-level acquisition configuration and control software architecture. Detailed description in the text.

During top-level acquisition configuration and control script execution, the constructs describing the TX sequence are translated into a bitstream (a set of register writes) that is loaded to the TXPB-256 hardware. In the case of acquisitions typical in 2-D SWE mode, a set of templates was developed as a part of the HAL driver to configure common methods like SSI or CUSE. In those cases, the user provides only high-level parameters related to pushing beam generation (frequency, duration, focal configurations, and others) and parameters related to the PWI tracking sequence (frequency, number of cycles, plane wave angles, and others).

In the future, the functionality provided by the HAL driver will be moved to ARRUS API (described below), but at the moment of writing this text, those are still separate software components, not integrated yet in a single package.

- **ARRUS API**

  The name ARRUS is an abbreviation for *API for Research Ultrasound*[1]. This API is a part of the software development kit (SDK) distributed with the us4R-lite system. It provides a set of tools to initialize the system, configuration of custom TX/RX sequences, mechanisms of data capture synchronization with acquisition execution, and many others. In this work, since custom TX hardware was used, only the RX configuration was done using ARRUS API calls. Moreover, ARRUS API comes with a framework that allows to integration of custom data processing blocks with the API to build a complete processing pipeline. Python version of ARRUS revision 0.8.0 was used in this work.

- **Data processing blocks libraries**

  For the purpose of custom data processing required for image reconstruction, a set of data processing function blocks was developed. The code of these functions was developed in Python and was integrated with a framework provided by ARRUS API for efficient integration with the us4R-lite scanner. A detailed description of these routines is given further in this chapter. In addition, a set of Python

---

[1]ARRUS is an open source software distributed under MIT license. It is supported by us4us Ltd. and available under link: https://github.com/us4useu/arrus

data analysis and presentation procedures was developed as auxiliary tools used for experiments described in the next chapter of this dissertation.

- **Top-level acquisition configuration and control script**

  A top-level Python script uses all the above components to bring the whole acquisition and data processing pipeline configuration into one place. It configures the us4R-lite with TXP-256 transmit extension module TX/RX sequence according to a set of high-level parameters provided by the user. In addition to the TX configuration done using the TXPB-256 HAL driver (described above), the script configured the RX sequence, including the number of receive frames, data acquisition time per shot, receive apertures definitions, and others. The script also initialized the system, controlled the execution of pre-defined acquisition, and instantiated the data processing pipeline. Finally, the script saved the captured data (processed and/or raw data) on the disk to be analyzed later.

  There were a number of top-level scripts developed, depending on the intended usage:

  - scripts automating specific experiments or measurements,
  - scripts configuring real-time imaging like B-mode imaging, or various 2-D SWE acquisitions incorporating various push sequences or shear wave tracking strategies.

3.3   Image reconstruction algorithm

For the purpose of stiffness map reconstruction using the acquired data, a dedicated multi-step algorithm was developed in two environments. An early prototype was developed in Matlab (The Mathworks Inc., Natick, MA, USA). This implementation allowed off-line image reconstruction using frames of raw channel RF data as the input. Then, a validated algorithm was moved to Python and adopted for real-time execution in GPU-accelerated implementation. This process is described in section 3.4. Most of the data processing concepts used in the developed algorithm were described earlier in this thesis — please refer section 2.3 for details. Therefore, a brief overview of the algorithm is provided below, explaining the selection of methods for each processing step and providing additional details wherever uncommon approaches were introduced. One of the target features of the reconstruction algorithm implementation was to provide high flexibility. Therefore, the algorithm has many configurable parameters to, for example, allow evaluation of their effect on the final image quality and computation time. Those parameters with their default values are also listed in the description below.

1. **Loading data**

   In the first step, the algorithm loads the frames of raw RF channel data dumped by the usR-lite system to the host memory after or during acquisition.

2. **RF data filtering**

   To improve the SNR of the input data, a symmetric pass-band Finite Impulse Response Filter (FIR) was applied to filter out frequency components coming from electronic noise.

Configurable parameters: number of filter coefficients (236), filter cut-off frequencies (4–7 MHz), and passband to stopband transition width (1 MHz).

3. **Demodulation**

Quadrature demodulation was used to bring the signal to the baseband. In-phase and quadrature components of I/Q data were then filtered by a low-pass FIR filter to cut off the demodulation products at the doubled fundamental frequency (f_c).

Configurable parameters: demodulation frequency (f_c), number of FIR filters coefficients (128), FIR filters low-pass frequency ($0.5f_c$), and FIR filters passband to stopband transition width ($0.5f_c$).

4. **Beamforming**

The implementation of a standard DAS beamforming algorithm for PWI was used. By default, the reconstruction was done on the grid with a pixel size of 0.2 x 0.2 mm with dynamic focusing. Implementation of this algorithm was taken from the ARRUS package developed by us4us Ltd. The original author of this part is Piotr Karwat (us4us Ltd.). This applies to both Matlab implementation and a kernel used in real-time application. After beamforming each frame, a walking average approach was utilized to perform angle compounding of the beamformed data stage as in the coherent plane-wave compounding (CPWC) method.

Configurable parameters: grid size (38 mm and 42 mm in lateral and axial directions, respectively), grid step (0.2 mm), transducer pith (0.298 mm), speed of sound (1540 m/s), transducer elements lateral directivity range for dynamic focusing ($\pm30^o$), receive aperture apodization (Hanning window).

5. **Shear wave motion detection**

Since I/Q data were available after beamforming a phase shift method was used to obtain shear wave motion data. Both Kasai's 1D-autocorrelator and Loupas' 2D-autocorrelator algorithms were implemented (see section 2.3.2). By default, Kasai's method in progressive referencing was used, axial kernel size of 0.8 mm and ensemble length of 4 frames, unless otherwise mentioned.

Configurable parameters: axial kernel size (0.8 mm), ensemble length (4 frames).

6. **Motion data filtering**

A directional filter (as described in section 2.3.3) was applied in (k, ω) space to separate leftward and rightward propagating waves. In addition, the mask was designed to remove all the signal components of frequencies below 40 Hz and above 800 Hz and to filter out all data components related to shear waves propagating speeds out of range of interest. This range was set with respect to the expected target stiffness range within the ROI. The mask edges were apodized to avoid oscillations in the spatio-temporal domain. The 2-D FFT and inverse 2-D FFT algorithms were used to transfer data between spatio-temporal and (k, ω) domains. Before that, datasets were zero-padded in both dimensions (spatial dimension corresponding to axial depth, and time dimension corresponding to slow-time) to a number of samples being the nearest power of two. As a result of directional filtering, two datasets were constructed — related to leftwards and rightwards propagating waves.

Configurable parameters: frequency domain filter passband frequencies (40–800 Hz), SWS domain filter cut-off values (0.5–4.5 m/s), 2-D FFT lengths (input datasets zero-padded in both dimensions to the nearest power of two by default).

7. **Shear wave speed estimation**

A correlation-based time-of-flight estimation algorithm (as described in section 2.3.4) was implemented to obtain local SWS estimates. A normalized cross-correlation (NCC) of Tukey-windowed tissue motion profiles at laterally distant locations (distance defined by lateral kernel size) was calculated to find propagation delay. In order to speed up the process, instead of calculating the NCC of the profiles as defined in Eq. (2.4), the correlations were found using FFT approach. The correlation theorem says, that the multiplication of the Fourier transform $F(u)$ of one function $(f(t))$ by a complex conjugate of the Fourier transform $G(u)^*$ of the other one $(g(t))$ gives as a result the Fourier transform of their correlation R_{fg}:

$$R_{fg} = f(t) \otimes g(t) \leftrightarrow F(u) \cdot G(u)^*, \tag{3.2}$$

where the $\otimes$ symbol represents correlation. To apply this approach, first, both particle velocity profiles were taken to the frequency domain using FFT. Then the complex conjugate of one of the signals was found and both results were multiplied together. Finally, the inverse FFT was performed. Normalization of the results was also done so that the autocorrelations at zero lags were equal to 1:

$$R_{fg,n}(j) = \frac{1}{\sqrt{R_{ff}(0) \cdot R_{gg}(0)}} \cdot R_{fg}(j), \tag{3.3}$$

where j is the correlation lag, $R_{ff}(0)$ and $R_{gg}(0)$ are autocorrelations at zero lags. The above calculations were performed for each target image pixel. Before that, shear wave motion data were interpolated by a factor of 5 (by default) along the slow-time dimension using the spline interpolation method of order 2. The above algorithm was applied separately on two input datasets coming from the directional filtering step, producing for each an SWS map and a map of correlation coefficients that was treated as the estimation quality indicator map.

Configurable parameters: Tukey window cosine fraction (shaping the Tukey window, 0.1 by default), interpolation factor (5), spline interpolation order (2), and lateral kernel size (2.8 mm in homogeneous experiments, 4 mm in homogeneous experiments).

8. **Image compounding**

The approach as described in section 2.3.5 was used to combine multiple SWS maps to build a single compounded SWS map. Mask boundaries were set experimentally, depending on the pushing sequence design. In the regions where obtained SWS maps overlapped the SWS estimates were calculated as the weighted sum, taking the correlation coefficients as weights. To achieve smooth transitions between data from different input SWS maps, a sigmoid-based data weighting across the region boundaries was applied.

Configurable parameters: mask boundaries (depend on pushing beam design).

9. **Post-processing**

   In order to refine the image, median filtering was applied to the SWS map. Kernel size of the filter of 1 mm x 1 mm was used unless otherwise mentioned.

   Configurable parameters: median filter kernel dimensions (1 mm x 1 mm).

   In 2-D SWE mode scenarios involving more than one acquisition to reconstruct a single image, like in the SSI method, steps 1–7 were performed for data acquired from each acquisition. Then, a single SWS map was constructed in step 8, and step 9 was executed normally.

3.4 Real-time imaging implementation

Although in experimental setups the approach involving off-line image reconstruction after capturing subsets of data may be sufficient, clinical applications typically demand real-time imaging. This requires rapid data acquisition and processing to minimize the system latency and achieve a frame rate that — combined — provide instantaneous feedback. Such feedback plays a critical role in the precise targeting of the desired area and achieving the desired image quality. This is accomplished by utilizing the resulting image as a guide for probe placement. Anatomical B-mode imaging or Doppler-based blood flow imaging are examples of ultrasound modalities that benefit from a real-time operation, reaching frame rates in the range of tens of frames per second (FPS). The benefits of real-time stiffness estimates capture were also reported in SWE technique [181]–[183].

In 2-D SWE, even though elevated frame rates are both desirable and relevant from a clinical perspective, fundamental limitations are imposed by thermal safety concerns related to the repetitive generation of push pulses. As described earlier in section 2.4, a repetitive SWE data acquisition performed with a high duty cycle may result in undesired tissue thermal effects or excessive transducer heating. A systematic experimental and modeling study held by Fahey et al. [150] has shown that due to thermal effects, the frame rate should be kept under 0.4 FPS over a 2-minute window in conventional ARFI imaging utilizing 40 push beams of length 320 μs to build a single image. It was also concluded, that if a greater number of parallel tracking channels were available, higher frame rates would likely be safely achievable. This is the case in the 2-D SWE method which utilizes high-frame rate plane wave imaging for shear wave tracking; it does not require that many pushing beams and acquisitions to build a single image. Nevertheless, based on this information and assuming that only 10 push beams would be required to form a single image, frame rates below 2 FPS would be achievable. It is still a rough estimation, since, for example, parameters related to push generation and acquisition may be adjusted to maximize frame rates by sacrificing the image quality. System operation at high frame rates would also consume more power, which can be a limiting factor in portable scanners. To summarize, a careful system calibration should be performed, since the significance of these thermal effects may be system- and application-specific.

Since the reconstruction algorithm is more complex than used in conventional B-mode or Doppler imaging, another factor potentially limiting the maximum imaging frame rate is the system's computational performance. This topic is explored in the next section.

3.4.1 Performance requirements

Even at relatively low frame rates of several FPS, as limited by the thermal constraints discussed above, providing sufficient data processing throughput for the 2-D SWE method can be a challenging task in a system design. One must ensure the complete data processing pipeline operation in a strict timing regime for real-time imaging. In a software-based imaging platform such as us4R-lite, the whole pipeline typically consists of 3 main stages:

1. raw echo RF data capture,

2. data transfer to the computing back-end (PC),

3. data processing and image display.

For the purpose of estimating of required data transfer throughput and computational performance in the 2-D SWE method, the following assumptions are taken:

- data capture from a 128-element probe and ADCs operating at 65 MHz RF sampling rate (with 16 bits per sample),

- data acquisition sequence as in the SSI method: 3 acquisitions to reconstruct a single image,

- CPWC imaging for shear wave tracking to an axial depth of 42 mm at PRF of 5 kHz,

- shear waves tracking for 20 ms in each acquisition,

- the algorithm as described in section 3.3 with a default set of reconstruction parameters is used. The following values of configurable parameters are used.

A)   Memory bandwidth requirements

In the scenario specified above, each TX/RX event would generate ∼3600 samples in each receive channel. It gives in total raw receive RF data frame of size equal to ∼900 KB. With the PRF of 5 kHz and the tracking sequence lasting 20 ms, the data acquisition process would produce ∼100 frames of the volume of ∼88 MB. Since to reconstruct a single image 3 acquisitions were assumed, 88 MB x 3 = 264 MB of raw RF data would need to be transferred to the PC for further processing for each image. As noted in section 3.1.1, the us4R-lite system transfers data to the PC over the Thunderbolt interface with a theoretical bandwidth of 2.4 GB/s. This means, that the interface bandwidth would be the bottleneck for frame rates higher than ∼9.5 fps.

B)   Computational performance requirements

In applications like 2-D SWE implemented using a software-based system, it often turns out that the computational performance of the processing back-end becomes a limiting factor for achieving higher frame rates. Regarding incoming data traffic being in the order of hundreds of megabytes every second, the processing back-end shall be equipped with sufficient computing resources to handle the processing of such a large input data volume. In order to roughly estimate the computational workload of the image reconstruction in the 2-D SWE method, the developed algorithm's code was analyzed to

Table 3.4: Number of floating point operations per each algorithm's step (approx.) required to reconstruct data from a single acquisition.

Step	Number of operations ($\cdot 10^9$)	% of total
RF data filtering	26.9	16%
Demodulation	23.6	14%
Beamforming	8.7	5.2%
Shear wave motion detection	0.8	0.5%
Motion data filtering	14.7	8.7%
SWS estimation	93.6	55.5%
Post-processing	0.2	0.1%
Total	**168.4**	**100%**

extract a number of floating point arithmetic operations required per each processing step. It is summarized in Table 3.4.

This simplified analysis shows that the SWS estimation requires most of the computational power of all processing steps performing $93.6 \cdot 10^9$ arithmetic operations, which is 55.5% of the total, that is $168.4 \cdot 10^9$ operations to reconstruct an image from a single acquisition. Since it was assumed that 3 acquisitions are required to build a final image, $\sim 505 \cdot 10^9$ floating point operations would be needed in total. Therefore, in order to achieve a frame rate of 1 fps, one would need to provide a computational performance of a processing unit equal to ~505 GFLOPS (giga (or billion) floating point operations per second). For reference, the MSI GS66 mobile workstation used as a host in the experimental part of this thesis features the 10th generation Intel Core i9 10980HK CPU with 18 cores, which performance is benchmarked by the manufacturer to 307.2 GFLOPS [184]. This would allow to achieve 1.8 fps at best, assuming constant CPU operation at peak performance, which is not realistic.

It should be noted, that the performance requirements estimation shown above has a number of limitations. First of all, the estimated number of floating point arithmetic operations required per step is done very roughly. In practice, to perform scientific calculations, specialized libraries can be used, which may optimize the code for improved performance. Moreover, each type of arithmetic operation was treated in the same way, whereas in reality, they may significantly differ in the number of required instructions, e.g. addition and square root. Furthermore, execution of the algorithm's code involves manipulation on data that are not arithmetic operations but may bring significant overhead to total processing time, e.g. data transfers or data access. However, rather than accurately estimating the required computational capability, the aim of this analysis was to obtain some means of algorithm complexity to pick an efficient way of implementation.

Initially, the algorithm code was developed in the Matlab environment. This environment, widely used in the scientific community, provides a wide range of ready-to-use functions and tools that allow convenient and fast development of complex data processing algorithms. This implementation allowed off-line image reconstruction and was used to tune configurable parameters and perform early system validation. The complete algorithm execution in Matlab took 1–2 minutes, depending on parameter values, which was good enough for testing, but it was not even close to real-time performance.

The us4us Ltd. company members, including the author of this dissertation, prepared a short-course titled *"Ultrasound Signal Processing with GPUs — Introduction to*

Parallel Programming" that was held during the 2022 International Ultrasonics Symposium in Venice, Italy. For the purpose of this course, the simplified and limited Matlab version of the 2-D SWE image reconstruction algorithm was adopted to execute on GPU and was presented as a case study. With minimum effort, the GPU-accelerated version was developed and it outperformed the Matlab version in terms of execution time by nearly three orders of magnitude. At this point, it became clear, that the real-time performance of the algorithm was within reach and should be achievable with further optimizations for GPU execution.

3.4.2 General purpose processing using GPUs

Until 2003, microprocessors based on a single-core central processing unit (CPU) exhibited rapid performance increases between each chip generation. This drive allowed to improve the execution speed of software at grown application complexity. However, this trend has slowed down since 2003 due to energy consumption and heat dissipation issues at the silicon level that have limited the increase of CPU clock frequency [185]. To address this issue, processor manufacturers switched to architectures that relied on multiple processing units (multi-core), in which multiple threads of execution could co-operate to complete faster. First multi-core processors featured two cores while current state-of-art general-purpose processors consist of up to 16 cores, each running with a clock frequency reaching 5 GHz.

In contrast, rather than optimizing the execution of a sequence of operations as in multiple-core CPUs, the GPU architecture focuses on the optimization of parallel processing in thousands of threads executed by a large number of smaller multi-threaded cores, amortizing this way the slower single-thread performance to achieve greater throughput. The difference in these two devices' architectures is illustrated in Fig. 3.8. This approach allowed GPUs to lead the computational performance race since 2003 [185] and the performance gap between GPUs and multi-core CPUs has been increasing until today. Nevertheless, it should be noted, that since GPUs are designed to accelerate heavily parallel numeric calculations, they will not perform as well as CPUs in tasks they are optimized for. Due to these complementary attributes of CPUs and GPUs, heterogeneous CPU/GPU co-processing systems evolved that allow applications to perform best using both types of processors, each for different tasks [186].

A) CUDA programming model

Initially, programming the GPUs for general-purpose applications — other than graphics processing — was very difficult and inconvenient, because GPUs had fixed architecture for graphics processing and programmers had to use graphic application programming interface (API) functions to access the GPU resources. It required to define a general problem in a way that followed graphics processing principles. This way graphic APIs along with limited a range of applications that were suitable for GPU.

The way of programming the GPUs changed significantly in 2006 with the release of the first GPU based on CUDA architecture (Compute Unified Device Architecture) [188]. GPUs based on this architecture were designed specifically to overcome limitations that used to make GPUs not suitable for many general-purpose applications. With CUDA, using graphics API was no longer required and the programmer was granted direct access to GPU's computational resources. CUDA is a hardware and software coprocessing architecture for parallel computing that enables NVidia GPUs to execute programs
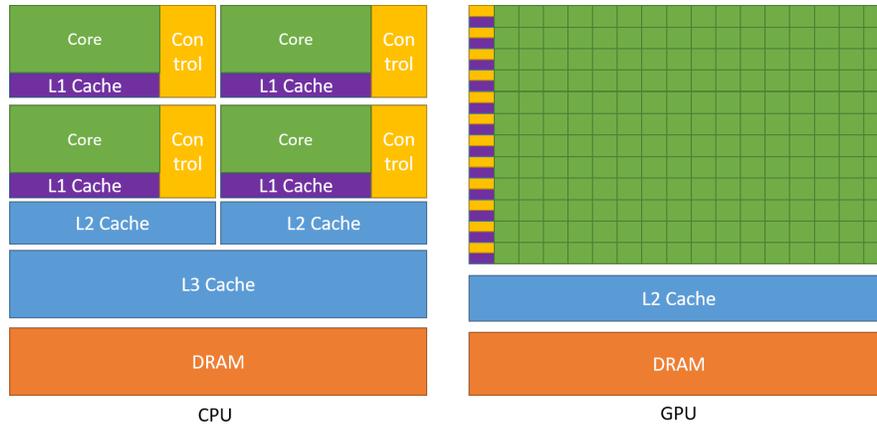
Figure 3.8: An example distribution of chip resources for a quad-core CPU and a GPU. The GPU is specialized for massively parallel processing which is reflected in its architecture. It is designed in a way that more resources are devoted to data processing rather than data caching from memory and flow control for minimized latency. DRAM stands for Dynamic Random Access Memory. Image source: [187].

written with commonly used programming languages, like C or C++. Because most languages were designed for one sequential thread, CUDA provides a minimalist set of extensions for expressing parallelism, e.g., CUDA C.

Typically, a CUDA program is organized into a host code, consisting of a sequential thread running on a host CPU, in which functions called kernels are called to be run on GPU (see Fig. 3.9. Kernels are executed N times in parallel by N different CUDA threads. This way, the application performance can be boosted due to data parallelism — while threads execute the same code, each operates on different input data. The phases that exhibit little or no data parallelism are implemented in the host CPU code. The expected order of overall speed-up possible to achieve depends on the portion of the application that can be parallelized. It will be shown later, that the 2-D SWE image reconstruction algorithm features significant data parallelism.

B) CUDA hardware architecture

In the typical CUDA architecture, the GPU is organized as an array of multi-threaded streaming multiprocessors (SMs). The number of available SMs varies between GPU generations. The parallel processor cores (CUDA cores) within each SM share control logic and execute instructions for parallel threads. By exploiting multithreading, multiple SMs provide scalable data and task parallelism to execute multiple thread blocks in parallel, and the number of parallel threads is typically counted in thousands [185]. In fact, efficient threading support allows applications to expose a much larger amount of parallelism than available hardware execution cores with little or no penalty [186]. The work scheduler distributes CUDA thread blocks to SMs with available capacity, dynamically balancing the computing workload across the whole GPU. Such thread execution control helps to limit the impact of long latency accesses to DRAM which is known to be crucial in achieving high performance [158], [185]–[187]. Usually, however, dedicated measures need to be developed to avoid execution overheads due to excessive memory transfers. In general, efficient memory management can be facilitated through methodical use of GPU's memory structure, which consists of (see Fig. 3.10): shared memory

77
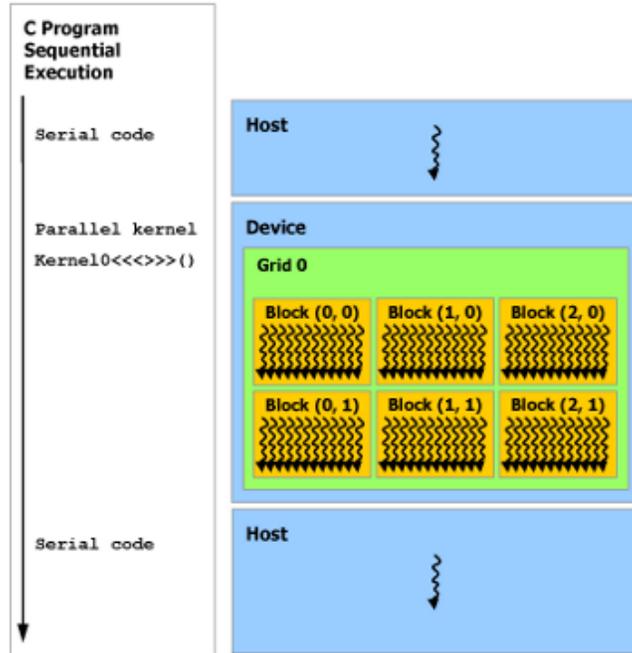
Figure 3.9: CUDA heterogeneous programming model. A grid of parallel threads is organized as an array of blocks, which in turn consists of an array of threads. Image source: [187].

for each thread block (small size, fast access); texture and global memory (slower access speed, large size). Known strategies are described in literature [185]–[187].

In this thesis, the NVidia GeForce RTX 3060 GPU was used as an accelerator for the real-time 2-D SWE image reconstruction algorithm. Table 3.5 lists a number of available resources and selected parameters of this GPU. Importantly, the maximum 32-bit floating point performance is much higher than the required performance estimated for SWE reconstruction in the previous section. It should be noted, however, that this is not necessarily the achievable speed of real application, but rather the raw speed that the execution resources can potentially reach. Frequently, traffic congestion in the GPU's global memory access paths limits the performance of the GPU-accelerated application and only a small fraction of the potential speed of underlying hardware can be achieved without code optimization to reduce memory traffic [185], [186].

C)   *Software scalability*

Scalability is another important feature of GPU processing using CUDA. In CUDA, the computational problem decomposition allows threads to cooperate when solving each sub-problem, and at the same time enables automatic scalability — each block of threads can be scheduled on any of the available SM within a GPU, in any order, concurrently or sequentially, so that a compiled CUDA program can execute on any number of SMs. This makes the software design easily transferable between GPU devices with a different number of resources. For example, an application accelerated by a high-power, high-performance GPU can be transferred to a low-power, mobile GPU with minimum effort.
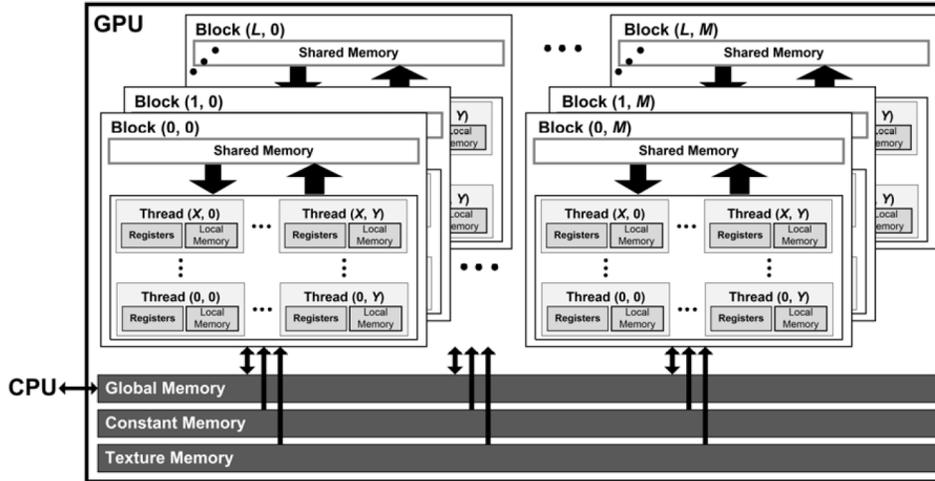
Figure 3.10: In CUDA-capable GPU, host and device have separate memory spaces. The input data needs to be transferred to the device memory to execute a kernel. During kernel execution, on-chip shared memory accesses are preferred due to its lower latency as a result of its higher memory bandwidth (arrows width are indicative of bandwidth). Source: [163].

Table 3.5: Specification of NVidia GeForce RTX 3060.

Specification	Value
Architecture	Ampere
Base clock speed	1.32 GHz
SMs	30
CUDA Cores	3,584
Memory	8 GB GDDR6
Memory bandwidth	360 GB/s
L1 cache (per SM)	128 KB
L2 cache	3 MB
FP32 (float) performance	12.74 TFLOPS
FP64 (double) performance	199 GFLOPS

3.4.3 Data parallelism in 2-D SWE image reconstruction

It was mentioned in the previous section, that the expected application speed-up depends largely on the portion of operations that can be performed in parallel. The higher data parallelism is present in the computational-intensive problem, the higher speed-ups should be expected. In order to predict the theoretical speed-up when using parallel processing Amdahl's law is often used:

$$S_{latency}(s) = \frac{1}{(1-p) + \frac{p}{s}}, \tag{3.4}$$

where $S_{latency}$ is the theoretical speed-up of the execution of the whole task, s is the speed-up of the part that benefits from improved system resources (i.e. parallel processing), and p is the proportion of execution time that the improved (parallelized) part originally occupied. Analysis of the formula above for the case of $s \rightarrow \infty$ shows that although the theoretical speedup increases with the improvement of the resources of the

system and regardless of the magnitude of the improvement, the theoretical speedup is always limited by the part of the workload that cannot benefit from the improvement.

A brief analysis of the reconstruction algorithm for the 2-D SWE method, as described in section 3.3 shows that each step of the algorithm exhibits massive data parallelism:

- RF data filtering and demodulation — each input data frame, each waveform within a frame, and each sample within an output waveform can be processed independently, in parallel.

- Beamforming — each pixel within each output frame can be processed in parallel.

- Shear wave detection — each pixel within each output frame can be processed in parallel.

- Motion data filtering — each axial slice of three-dimensional motion data can be processed in parallel and each element of the slice in (k, ω) space can be processed in parallel. The same applies to masking and inverse transform.

- Shear wave speed estimation — each target pixel can be processed in parallel and cross-correlations for each lag can be found in parallel.

- Post-processing — each target pixel can be processed in parallel.

In addition, if the final image is formed from more than one acquisition, data gathered from each acquisition can be processed in parallel before final compounding. The above analysis makes it reasonable to expect significant speed-up in application execution by utilizing parallel processing techniques using GPU. However, although the problem shows data parallelism at many levels, an efficient implementation should consider underlying hardware execution to choose the proper way of parallelization and computational resources assignment. Examples of such considerations are presented in section 3.4.5. It should be also noted, that each processing step can be run only after the previous step has been completed providing input data for the next step.

3.4.4 Software development tools

This section presents a set of tools and methods used for the development of real-time 2-D SWE image reconstruction application.

Running a CUDA application requires the system with at least one CUDA-capable GPU and a device driver that is compatible with the *CUDA Toolkit*. *CUDA Toolkit* provides an environment for the development and deployment of high-performance GPU-accelerated applications: a compiler, development tools, GPU-accelerated libraries, a debugger, a profiler, and the CUDA runtime. All those tools are distributed by NVidia out of charge and can be downloaded from its website [189]. For the application described below *CUDA Toolkit* version 11.8 was used.

The host application code was developed in Python programming language version 3.8. Since the arithmetic operations on large data constructs such as arrays tend to execute slowly in the interpreted Python code, *NumPy* library was utilized to speed up the calculations. The *NumPy* (Numerical Python) is an open-source Python scientific computing library that is used in many fields of science and engineering. The *NumPy* library contains multidimensional array and matrix data structures with methods to efficiently

operate on them. Along with efficient implementation of common arithmetic operations with arrays and matrices, it supplies a library with a broad portfolio of high-level mathematical functions that operate on these arrays and matrices. More information on *NumPy* can be found on its webpage ([190]). Another library used in the developed application was *SciPy*. The *SciPy* is a collection of mathematical algorithms and convenience functions built on the *NumPy*. It contains efficient implementation of many well-known algorithms for interpolation, digital signal processing, image processing, and many others. More information on *SciPy* can be found on its webpage ([191]).

Although the two library packages described above were found useful and effective in the implementation of numerical calculations for SWE method image reconstruction, they give no support to moving the execution of computational-intensive parts of the application from CPU to GPU. For this purpose, the CuPy package was chosen. The CuPy is *NumPy/SciPy*-compatible open-source array library for GPU-accelerated computing with Python. It utilizes *CUDA Toolkit* libraries to make full use of the GPU architecture. In the most basic way, CuPy can be used as a drop-in replacement for *NumPy* and *SciPy*, since it brings GPU implementation of many functions of these libraries. This simple approach was reported to achieve overall performance growth in a range from 10% to 2700x, depending on the operation type and data size [192]. Additionally, CuPy allows straightforward integration of custom CUDA kernels developed in CUDA C. In the developed application efficient GPU implementations of commonly used functions available in CuPy were combined with custom kernels that implemented data processing that required careful optimization. More information about CuPy can be found on its home page ([193]).

3.4.5 GPU-accelerated reconstruction algorithm implementation

This section describes in detail the real-time implementation of the 2-D SWE image reconstruction using GPU. A top-level software architecture description is followed by a detailed elaboration on each processing kernel implementation.

3.4.5.1 Top-level architecture

The software architecture takes advantage of tight coupling between the us4R-lite system and a PC equipped with a GPU to build a streaming processing system. Fig. 3.11 conceptually illustrates the data flow within the software. Firstly, before any acquisition is started, the imaging platform is loaded with a TX/RX sequence configuration that is found based on top-level parameters specified by the user. This process was described in section 3.2. After the acquisition, raw RF data collected by the us4R-lite are streamed to the PC RAM. Then, the PC is responsible for the data processing to reconstruct images in real-time. For this purpose, the total processing workload was split between CPU and GPU. Most of the processing was implemented as a set of GPU computing kernels to leverage the massive data parallelism. Kernels were developed to implement most of the SWE image reconstruction steps, as shown in Fig. 3.11.

Before any processing on the GPU can happen, the input data must be passed from the CPU (often referred as *host*) to the GPU (referred as *device*). Since GPUs are unable to directly access the host's memory, the batch of raw input RF data frames was first transferred to and stored in the device's global memory (DRAM) before invoking processing kernels. The general CPU-GPU coprocessing procedure is performed according to the steps below:
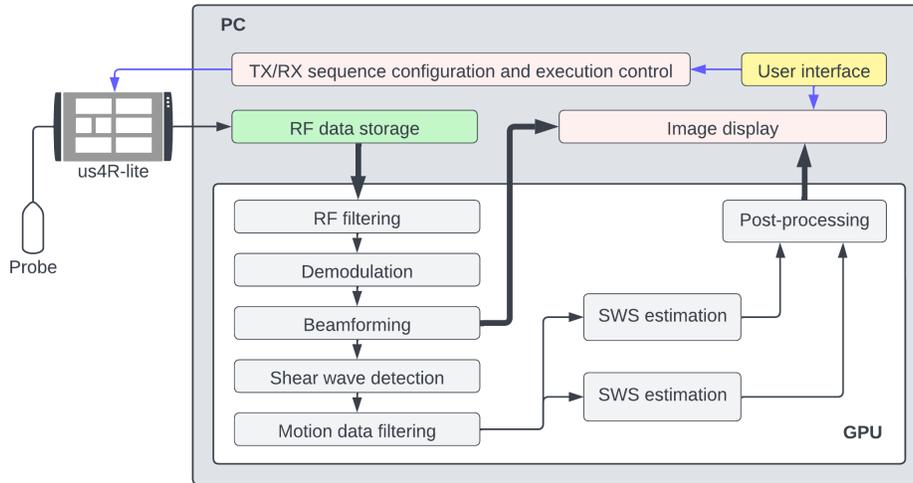
Figure 3.11: Top-level data flow architecture of GPU-accelerated reconstruction algorithm software. Blocks marked in light pink are executed solely on the CPU, while the gray blocks are implemented as kernels for execution on the GPU. The input data are streamed from the us4R-lite imaging system and transferred to the GPU for parallel processing in a series of steps. Motion data filtering step outputs two arrays related to rightwards and leftwards propagating waves and each is processed separately. Two types of data are taken for display purposes: beamformed data to display a B-mode image and a final stiffness map. For brevity, some intermediate processing steps were skipped in the diagram. For details, see section 3.4.5.2

.

1. Memory buffers for input and output data are allocated on the device memory and the input data is transferred from host to device memory space.

2. GPU kernel functions are invoked in the specified order to realize the algorithm. Operations are simultaneously executed over the grid of threads using the set of CUDA cores. Each kernel loads data from device's global memory and outputs the result array back.

3. Once GPU computation is completed, the output array is copied from the device memory back to the CPU host memory for display or optional storage.

In the presented application, two types of data arrays are taken for display purposes: beamformed data to display a B-mode image and a final stiffness map. Some final post-processing that precedes B-mode image display is performed on the CPU due to its negligible computational workload.

3.4.5.2 Data processing kernels

This section presents a description of each developed processing kernel implementation. Since each kernel performed a different type of calculations that exhibited different data access patterns, problem decomposition differed between kernels. Therefore, in addition to a general description of kernel functionality, some relevant kernel-specific considerations on performance are given in each section below.

A) RF data filtering

The RF data filtering kernel design is shown in Fig. 3.12. It followed the architecture as proposed by Yiu et al. in [99]. Assuming that the RF data is a 3-D array of dimensions $[M, N, P]$, where M is a number of channels, N is a number of samples, and P is a number of frames, the problem was decomposed as follows. Each thread block within the two-dimensional grid of threads was assigned to process a single channel waveform from a single frame. Thus, the grid of threads had dimensions of $[M, P]$. The kernel performed L-tap FIR filtering by convolving the input waveform with the filter impulse response. Before that, the input waveform was loaded to the shared memory, and filter coefficients were loaded to the constant memory to minimize memory access overhead. Since in general, N is larger than the maximum number of threads within a threads block, the input was decomposed into M-sample segments, and each was processed by a separate thread. Finally, the outputs from different threads were then concatenated and stored in the global memory of the GPU to be further processed. An equivalent implementation of this RF FIR filter was also written in Python and GPU-accelerated using the CuPy library.
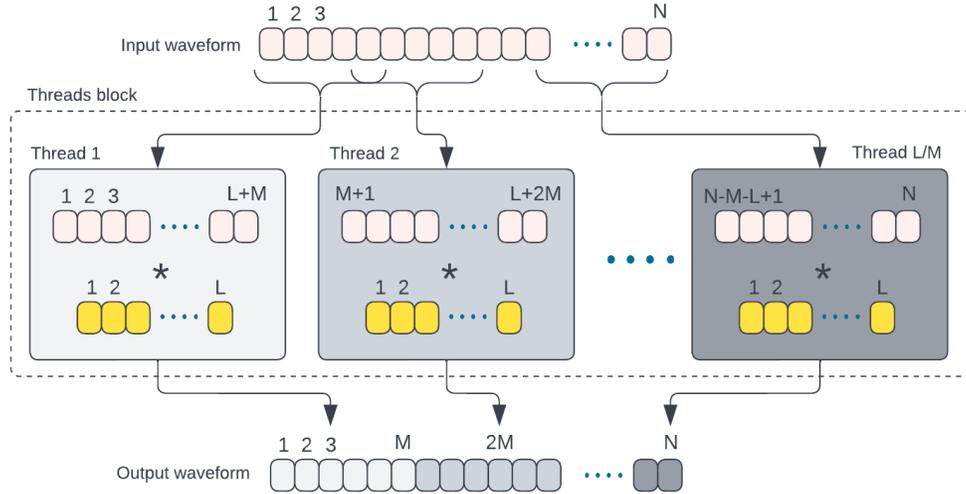
Figure 3.12: GPU FIR filter kernel design for RF raw echo signal low-pass filtering. The input waveform of length N is divided into M-sample segments (with an L-sample overlap). Each thread within a grid processes each segment by convolving the segment data with the filter coefficients. The "*" sign means convolution operation. Redrawn from [99].

B) Demodulation

The demodulation kernel used a similar way of decomposition as in the previous step. Kernel design is illustrated in Fig. 3.13. Whereas in the previous kernel, the filter coefficients were loaded, here the sine and cosine values were calculated within the threads. Then, the kernel performed quadrature demodulation — the samples of the given segment are multiplied by sine and cosine samples to obtain $I(n)$ and $Q(n)$ output signals. Outputs from different threads are then concatenated and stored in the global memory of the GPU to be further processed. An equivalent implementation of this demodulation kernel was also written in Python and GPU-accelerated using the CuPy library.
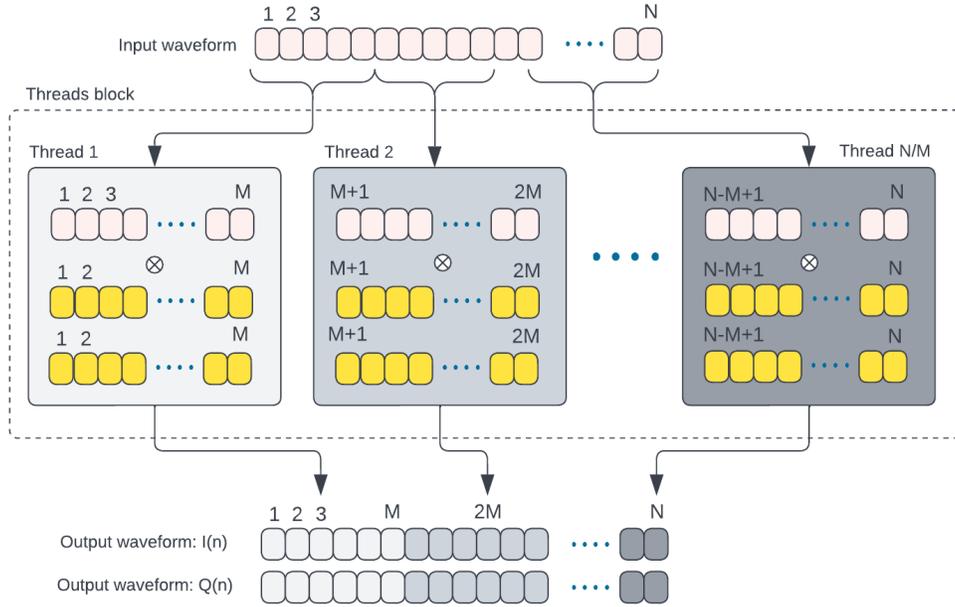
Figure 3.13: Demodulation kernel design for GPU execution. The input waveform of length N is divided into M-sample segments. Each thread within a grid processes each segment by multiplication of the segment values with the sine and cosine samples.

In this step, the outputs, namely $I(n)$ and $Q(n)$, are low-pass filtered using the same approach as presented in section A.

C)   Beamforming

The beamforming kernel design is shown in Fig. 3.14 and it follows the approach as proposed by Yiu et al. in [158] and then reused by the same group later in [99]. A pixel-based strategy was used here. Each thread within a grid of threads is assigned to find the beamformed value of one pixel in one image frame. A two-dimensional grid of threads covers the whole target image. Due to the expected locality of data, thread blocks were organized in a two-dimensional grid to cover segments of the target image, as illustrated in Fig. 3.14.

Regarding the operations performed by each thread, three main steps were involved to perform a conventional delay-and-sum (DAS) beamforming in the PWI method:

1. estimation of the focusing delays for all channels within the receive aperture with respect to the pixel position within the target image, which is equivalent to thread position within the grid of threads,

2. based on calculated delay, access to channel data values and perform linear interpolation to find sample value from each channel. It is found via linear interpolation of the two adjacent data samples that correspond most closely with the pixel's focusing delay in the given channel,

3. perform weighted summation of samples from each acquisition channel. Weight for the summation represents the receive apodization.

4. the output value is stored in the GPU global memory.

Compounded images were formed by coherently summing successive beamformed images in a walking sum method. For this task, another GPU kernel has been developed (not shown in Fig. 3.14). As mentioned earlier, this kernel was taken from ARRUS package and is originally authored by Piotr Karwat from the us4us company.
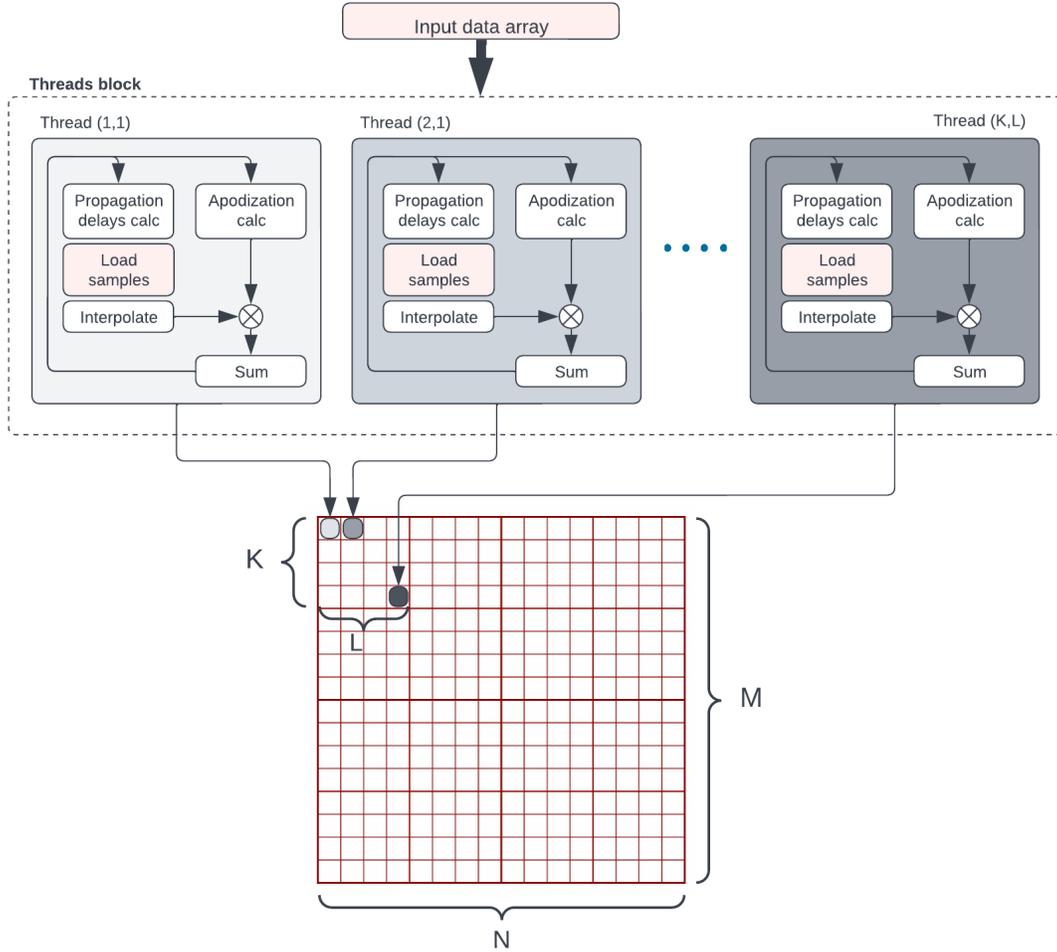
Figure 3.14: Pixel-based strategy for beamforming GPU kernel. Description in text.

D) Shear wave detection

Computational problem decomposition for parallel processing utilized in shear wave detection kernel used the same principle as described in beamforming kernel — the similar pixel-based processing was used. Each thread directly applied Eq. 2.2 formula to find local transient particle velocity. Since multiple neighboring threads accessed the same data samples, to improve memory access time, the input data were fetched into shared memory in each thread block. Shear wave detection reduced the amount of data by half since it operated on complex input data, but output real numbers.

E) Shear wave motion data filtering

In the shear wave motion data filtering step it was decided to use functions provided with the CuPy package, in particular: *cupyx.scipy.fft.fft2* and *cupyx.scipy.fft.ifft2* for direct

and inverse 2-D FFT. The performance of these implementations was found sufficient in this application.

F) Shear wave speed estimation

Similarly, as in the previous step, the implementation of shear wave speed estimation stage of the algorithm relied on functions provided with the CuPy package, in particular: *cupyx.scipy.fft.fft* and *cupyx.scipy.fft.ifft* for direct and inverse FFT and *cupyx.scipy.ndimage.zoom* for spline interpolation. It should be noted, that this step involved a reduction step, where multiple data frames were reduced to a single final image.

G) Post-processing - median filtering

This step was implemented using functions provided with the CuPy package, specifically, *cupyx.scipy.ndimage.median_filter*.

3.4.6   Real-time device operation

The developed implementation was a streaming processing system, where the acquired data were continuously transferred to the PC equipped with a GPU for image reconstruction as described in previous sections. This solution enabled obtaining of SWS maps in the real-time regime. Fig. 3.15 shows a snapshot from the display during on-going imaging. Three images are displayed — a B-mode image for reference and probe positioning, a reconstructed SWS map, and an animation made of reconstructed shear wave motion data. With frame rates operation higher than 1 fps the animation was replaced on display by a selected shear wave motion data frame. Achievable frame rates and processing performance topics are addressed in the next chapter.

In the real-time operation mode shown here, a custom push generation sequence was utilized, which combined principles of CUSE and SSI methods. Two parallel pushing beams were generated as in CUSE and the beams were focused at consecutive depths and generated sequentially as in the SSI method. Consecutive beams were focused at depths of 25 mm and 38 mm. This approach allowed to reconstruct SWS map within the whole FOV while keeping shear wave energy distributed over a high range of depth. It should be noted, that this is just one example out of many acquisition schemes possible to use for real-time 2-D SWE. Many other strategies can be prepared thanks to the high flexibility of the developed configuration software framework and image reconstruction algorithm.

Real-time operation of the presented system was presented during *2023 IEEE International Congress on Ultrasonics* in Montreal[2] and described in conference paper [194].

3.4.7   Data processing partitioning

In hybrid systems containing CPUs, FPGAs, and GPUs it is difficult to select an optimal data processing partitioning for sufficient computational performance, data throughput, high flexibility, and reasonable power consumption [195]. It is an especially challenging task in computational-intensive applications at a limited power budget, like in the case being the subject of this thesis. In the presented system design, efforts were put

---

[2]Video recorded for the online poster session can be found under the following link: https://vimeo.com/856770773
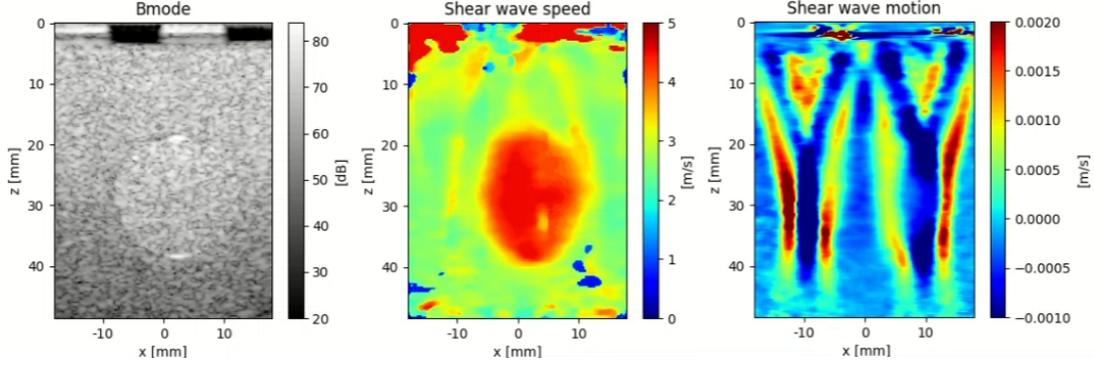
Figure 3.15: Images displayed during real-time 2-D SWE imaging using the developed system. Left: B-mode image. Middle: SWS map of the 70.9 kPa inclusion of 16 mm diameter. Artifacts are visible. Right: Shear wave motion data frame or animation.

to leverage the strengths of each type of processing device to find a balance between performance, configurability, and power. Table 3.6 gives the details on signal processing tasks breakdown by each device type.

Table 3.6: Signal processing partitioning per device type in the developed system.

Device type	Tasks
ASIC	- transmit beamforming, - analog signal processing: amplification with time gain control (TGC), pass-band filtering, - analog to digital conversion.
FPGA	- real-time TX/RX sequence execution control, - high-speed ADC data streams capture and buffering.
CPU	- acquisition planning, configuration, and scheduling, - data transfers to GPU and from GPU, - image display.
GPU	- raw RF channel data conditioning and down-conversion, - beamforming, - execution of SWE image reconstruction algorithm.

It should be noted, that signal processing conditioning as presented in Table 3.6 is a result of decisions made during specification of a software-based architecture of the us4R-lite system. In this architecture, the hardware processing part facilitated by ASICs and FPGAs is reduced in favor of increased processing within the software (CPU and GPU coprocessing). Hardware operation is practically limited to raw channel data capture and transferring it to the PC for further processing in software. The main advantage of this approach is that software solutions allow easy code modification and the development of novel methods.

Chapter 4

Experimental works

This chapter focuses on the experimental evaluation of the developed 2-D SWE imaging system. It consists of three parts. In the first part, the acoustic output of the system was measured to evaluate the safety of the imaging procedure. The second part describes the results of experimental works aimed at validating the developed system. In the third part, the system's processing performance is evaluated and the aspect of the system's power consumption is addressed.

4.1 Acoustic pressure measurements

As described in section 2.4, in medical applications it is required to characterize the acoustic output of the ultrasound device, not only for safety assessment or to evaluate the effectiveness of the procedure but also to meet the requirements of relevant regulatory bodies. This section presents measurements aimed at performing such characterization.

4.1.1 Methods

A dedicated test setup was built to measure the ultrasound pressure generated by the developed imaging system — see Fig. 4.1 for a detailed description. The 0.2 mm needle hydrophone (Precision Acoustics, Dorchester, United Kingdom) was used to perform the measurements. The pressure generated by two different pushing beams was measured — both of a center frequency 4.4 MHz and length of 300 μs, and using 120 Vpp excitation magnitude, but with different focal configurations:

- Push beam "A": focused at depth of 25 mm using 48 transducer elements of the probe (F/#=1.75),

- Push beam "B": focused at depth of 40 mm using 76 transducer elements of the probe (F/#=1.75).

The hydrophone sensitivity at the frequency of 4.4 MHz was declared to be 82 mV/MPa by the manufacturer at an uncertainty of 13%. The pressure field generated by the beams specified above was scanned in axial dimension along the beam axis and lateral dimension at the focus depth. The scans were performed in steps of 0.25 mm by shifting the probe position over the hydrophone using a manipulator. The hydrophone electrical response to acoustic pressure was recorded by the digital oscilloscope (RTB-2004, Rhode&Shwartz, Munich, Germany). Data collected this way was then used to derive
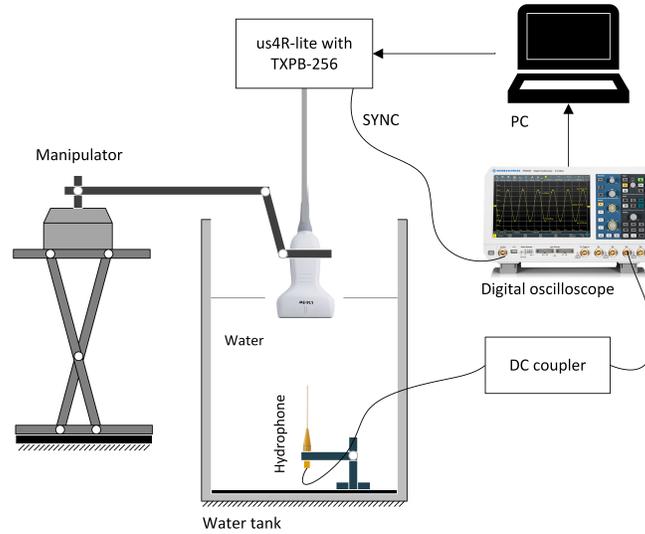
Figure 4.1: Acoustic pressure measurement setup. The needle hydrophone with submersible pre-amplifier was fixed in a custom holder placed at the bottom of the plexiglass water tank filled with degassed distilled water. An absorbing rubber plate was placed at the bottom of the water tank to reduce reflections. The output of the hydrophone was connected to the digital oscilloscope with 50 Ohm input impedance via a DC coupler. The ATL L7-4 probe was put in a waterproof ultrasonic probe cover and submerged in water above the hydrophone. The position of the probe was set by a manually controlled manipulator. The us4R-lite system with TXP-256 extension module was used to generate excitation for the probe. It also sourced the pulse for the oscilloscope to synchronize the data capture. Finally, the PC was used to configure the excitation of the us4R-lite system and analyze the data captured by the oscilloscope.

the ultrasonic field parameters as described in section 2.4, such as positive and negative peak pressures, peak-peak pressure, mechanical index, and derated spatial-peak temporal average intensity $I_{SPTA,0.3}$. The data capture was synchronized with the probe excitation moment by utilizing the us4R-lite system sync output. This also allowed to accurately measure the distance between the hydrophone tip and the probe by time-of-flight method. Since the pressure values obtained in hydrophone-based measurements are known to be highly sensitive to probe versus hydrophone alignment due to the directional response of the hydrophone [4], [147], [148], the setup positioning was carefully adjusted prior to the measurements using a short (4.4 MHz, 3 cycles), focused, repetitive pulse excitation.

4.1.2   Results and discussion

To expose some typical characteristics of the captured signals, an example waveform with the highest intensity measured for push beam "B" is shown in Fig. 4.2. The point of maximum intensity corresponded to the beam focal point. The waveform shows the beginning of the hydrophone response and the nonlinearity of the pressure waveform is revealed. This is a result of wave propagation through the nonlinear medium as is water [4]. While the peak positive pressure reached 5.3 MPa, the peak negative pressure was found much lower — −3.6 MPa.

The axial and lateral profiles of the pressure fields generated by the push beams "A" and "B" are plotted in Fig. 4.3. Push beam "A" generated higher pressure at the focal
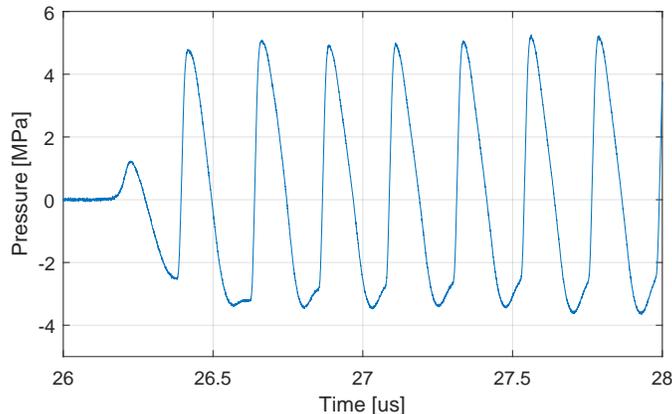
Figure 4.2: Hydrophone response for push beam "B" measured at beam focal point. The horizontal axis depicts the time elapsed from the start of transmission. Nonlinear waveform distortion is visible.

point than push beam "B" — 12.4 and 9.1 MPa peak-peak, respectively, even though push beam "B" used a higher number of transducer elements for transmission. This is mainly due to the greater depth of focus of push beam "B" and is a result of the effect of dissipation. Moreover, push beam "A" was focused at the depth of 25 mm, which is equal to the probe's elevational focus depth.

For each beam, maximum peak-peak pressure results differ slightly between scan types, e.g. for push beam "B" maximum pressure measured in the axial scan was 10.1 MPa, whereas in the lateral scan it was 9.1 MPa, while one would expect the same result. Although such a mismatch falls into the 13% uncertainty of the hydrophone, the real measurement error can be even higher. Even careful hydrophone-based pressure measurements are considered to have up to 20% uncertainty due to calibration errors, directivity, and spatial averaging [148]. Axial scans show a good agreement between desired beam's focus depths and the depths of maximum pressure. In the case of push beam "A", the maximum pressure was found at a depth of 25.06 mm, and 40.1 mm in the case of push beam "B". As expected, the lateral scans showed the maximum pressures at the beam's axes. This proves that the transmit delay profiles were applied correctly for desired focal configurations. The full width at half maximum (FWHM) for push beam "A" was measured to 1.5 mm, and 1.8 mm for push beam "B". Such difference is in the order of a scan's step, and considering measurement uncertainty, making any conclusions on this result would be premature.

Relevant parameters characterizing the acoustic output of the system in terms of safety are listed in Table 4.1. Those values were derived by hydrophone-based pressure measurements in water as described in section 2.4. I_{SPTA} values were obtained from the pressure waveforms captured at the locations of the highest value of acoustic intensity on the beam axis, which corresponded to the beams' focal points. Push PRF was assumed to 3 Hz. For reference, different derating values of 0.3, 0.5, and 0.7 dB/cm/MHz have been used to demonstrate what *in situ* intensities to expect in tissues with varying absorption levels.

In all cases, the measured quantities were below FDA limits. However, the obtained results suggest that caution should be taken when planning the TX/RX sequences for the SWE procedure. Regarding the mechanical index, push beam "A" showed a lower
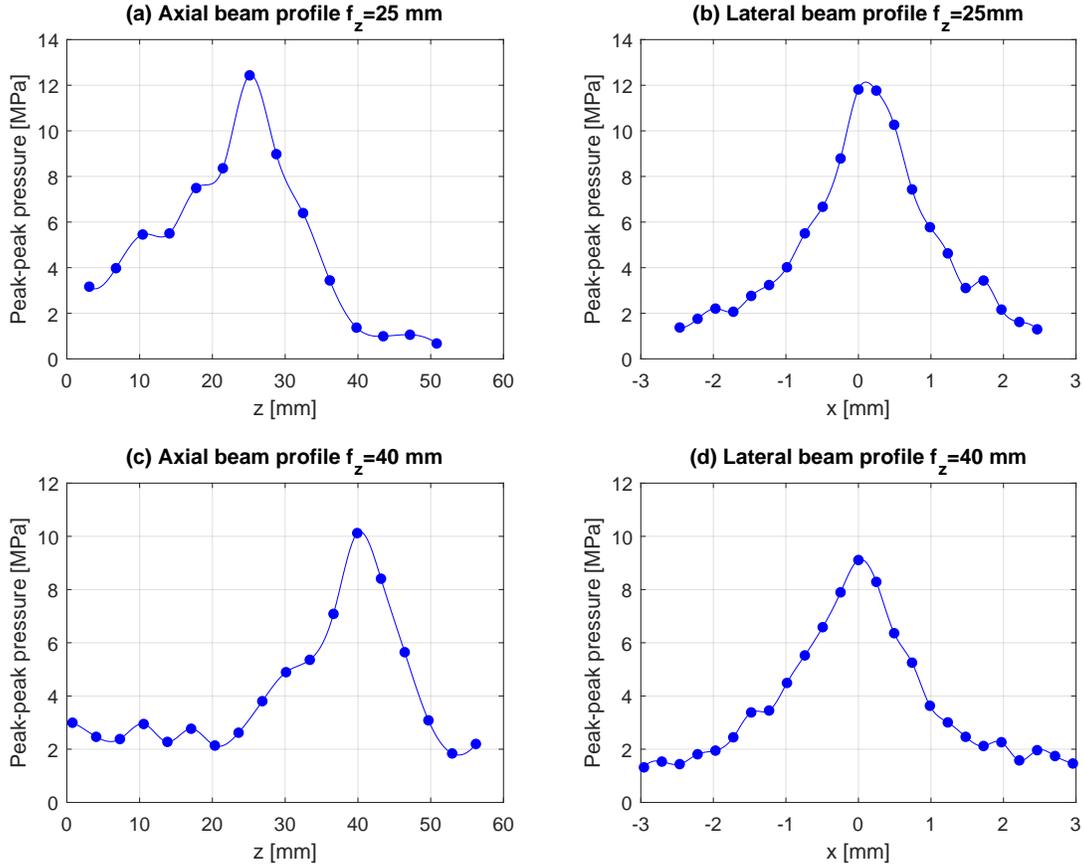
Figure 4.3: Peak-peak pressure profiles generated by evaluated push beams. (a): Axial profile of push beam "A". (b): Lateral profile of push beam "A", measured at a depth of focus (25 mm). (c): Axial profile of push beam "B". (d): Lateral profile of push beam "B", measured at a depth of focus (40 mm).

Table 4.1: Summary of measured acoustic field parameters compared with FDA regulatory limits. I_{SPTA} values were calculated assuming PRF equal 3 Hz. Subscripts in parameter names mean deration coefficient in dB/MHz·cm.

Parameter	Push beam A	Push beam B	Regulatory limit
$MI_{0.3}$	1.31	0.91	1.9
$MI_{0.5}$	1.01	0.58	N/A
$MI_{0.7}$	0.77	0.38	N/A
$I_{SPTA,\ 0.3}$ [mW/cm^2]	272	153	720
$I_{SPTA,\ 0.5}$ [mW/cm^2]	158	63	N/A
$I_{SPTA,\ 0.7}$ [mW/cm^2]	91	16	N/A

margin to the limit, reaching 1.31, which is 69% of regulatory 1.9, whereas push beam "B" reached 48% of the limit. Furthermore, those results were obtained using 120 Vpp excitation magnitude and f-number of 1.75. While increasing the excitation voltage or the number of active transmit elements is still possible within regulatory limits, it should be done carefully, especially at shallow focus depths. One possible strategy would be to make push pulse excitation voltage dependent on focus depth. Since, in general, less excitation energy is required to induce shear waves at shallower depths, one can use lower excitation voltage in this case to reduce the risk of exceeding the MI limit. However, dynamic switching of excitation voltage is not a common feature in ultrasound scanners.

Similar considerations should be conducted for I_{SPTA}. Although obtained results are much below the regulatory limits (maximum 272 mW/cm^2 measured at 720 mW/cm^2 limit), measured intensities were related to the generation of only a single pushing beam. Typically, however, 2D-SWE imaging sequence relies on a sequence of pushing beams in a pattern to induce shear waves in a big ROI, such as in SSI or CUSE methods as described in 2.1. Even though the pushing beams are focused at different points (in space and time), worst-case scenario analysis can be conducted in a similar way as proposed by Tanter et al. in [63]. In this approach, for *in-vivo* investigation, all pushing beams within a pattern can be assumed to be focused at the same location and resulting I_{SPTA} values can be summed together. For a possible scenario of using push beam "A" followed by a push beam "B" arranged as in SSI method, the total I_{SPTA} could be estimated to 425 mW/cm^2. In this worst-case analysis, combining such an SSI beam with another one generated simultaneously in an adjacent lateral location (as in CUSE method) would produce a total I_{SPTA} of 850 mW/cm^2 which exceeds FDA's limit. Again, same as in MI case considered above, the excitation plan can be modified to reduce the intensity levels. One simple solution would involve reducing push PRF, or reducing individual push pulse lengths or magnitudes.

The acoustic output study presented in this section has several limitations. First of all, as mentioned above, the presented measurement results suffer a significant level of uncertainty — up to 20%. Secondly, only a limited set of two push beam configurations were evaluated at fixed values of parameters such as excitation magnitude, push pulse duration, f-number, and frequency. Nevertheless, it was not the aim of this study to fully characterize the acoustic output of the system within a wide range of excitation parameters. It was conducted only to briefly check the safety of the procedure and give some means to coarsely evaluate possible safety margins of designed acquisition schemes. Finally, only the pushing beams' acoustic output was evaluated, completely omitting measurements related to the tracking sequence. However, is characterized by a very low I_{SPTA} value because it does not imply any focusing in the transmit mode [63] and it was considered negligible.

To summarize, obtained results have shown, that the acoustic output of evaluated push beams is within the regulatory limits. However, it was shown, that pushing patterns should be designed carefully, especially in real-time applications with high push PRF, since increasing the number of push beams or excitation strength in terms of push duration or magnitude can lead to exceeding the safety thresholds. Given that various imaging applications can have different pushing sequences, acoustic output safety characterization should be performed for each application. This study was conducted despite the fact that only phantom imaging is within this work's scope. However, obtained acoustic output measurements can be used to cross-compare the developed system performance to other implementations, since comparison based on excitation magnitude

can be misleading due to differences of used probes or efficiency of transmit subsystems.

4.2 System validation

In this part, a complete system consisting of the us4R-lite research platform with TXPB-256 extension module along with stiffness map reconstruction algorithm implementation described above were validated. The main goal of the validation was to check the overall system capability to acquire data in the 2-D SWE mode and to reconstruct meaningful stiffness maps out of this data. The operation of the system was verified experimentally, and the system performance of the shear wave speed measurement was evaluated by deriving chosen quality metrics of images acquired using industry-standard elastography phantom. Some intermediate results are shown to expose some typical problems appearing in SWE mode that may affect the imaging quality.

4.2.1 Methods

A) Push generation

All experiments were held using the 128-element ATL L7–4 linear array probe (Philips Healthcare, Amsterdam, The Netherlands) with a center frequency of 5 MHz and 0.298 mm pitch. For the validation purpose, the chosen push generation method followed the principle of well established SSI method, as proposed by the group of Tanter in [63]. There were 3 acquisitions used in total, each producing 3 laterally aligned pushing beams focused at successive depths generated sequentially and increasing the aperture size along with focusing depth to keep the f-number constant. Consecutive beams used 24, 50, and 66 elements and were focused at depths of 15 mm, 30 mm, and 40 mm, respectively (constant $F/\# = 2$). No aperture apodization was applied. Each individual push duration was 200 μs. The consecutive beams were generated with a pulse repetition interval (PRI) equal to 250 μs. See Fig. 2.2(b) for a reference. A push pulse frequency of 4.4 MHz was used in all experiments.

B) Data acquisition

A data acquisition scheme used in experiments was similar as introduced in section 2.2 (Fig. 2.3) and it is shown in the Fig. 4.4. The whole measurement procedure started with the generation of push beams. As described in the previous section, in SWE a push beam generation is followed by shear wave tracking. Immediately after the end of push generation, the us4R-lite scanner switched to high frame-rate PWI mode transmitting 2 cycles at a center frequency of 4.4 MHz with a pulse repetition frequency (PRF) of 10 kHz using all probe elements and capturing echo from up to 50 mm depth. In order to improve B-mode image quality, the CPWC method was utilized. A set of angles used was $[-4^o, 0^o, 4^o]$. In each acquisition, a total number of 150 frames were captured. Due to a limited number of 64 RX channels of the us4R-lite system, each transmission was repeated twice to capture the whole 128-element receive aperture, which reduced the effective frame rate twice, down to 5 kHz, and required 300 plane-wave transmissions per acquisition.
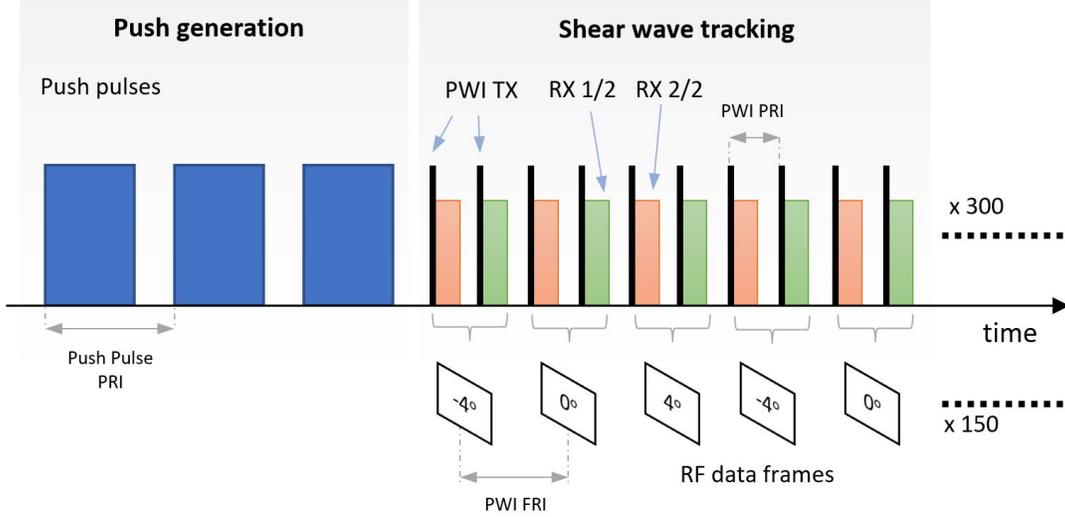
Figure 4.4: Example data acquisition scheme for SSI method. To capture RF data for a full receive aperture, two transmissions were required - each followed by 64 RX aperture size. PRI - pulse repetition interval, FRI - frame repetition interval.

C) Image reconstruction

The algorithm described in section 3.3 was used to obtain stiffness maps. At the shear wave speed estimation algorithm step, lateral kernel sizes of 3.2 mm and 5 mm were used in homogeneous and heterogeneous experiments, respectively. The rest of the reconstruction parameters were set to their default values, as specified in section 3.3.

D) Phantom experiments and image quality metrics

A complete test setup is presented in Fig. 4.5. An evaluation of the system performance was carried out using a tissue-mimicking Model 049A elasticity Q/A phantom (CIRS, Norfolk, VA, USA). The phantom had four stepped cylinder mass targets of known stiffness and six diameters. Targets of all available stiffnesses (types I–IV, nominal values of 4.7, 8.6, 34.0, and 70.9 kPa, or 1.25, 1.69, 3.37, and 4.86 m/s, respectively, if expressed in shear wave speeds using formula 1.19) and three diameters (6.5, 10.4, and 16.7 mm), all centered at depth of 30 mm were used. The phantom region with no inclusions (background) of nominal stiffness of 16.7 kPa (2.36 m/s) was used in the homogeneous region imaging experiment. Before performing the measurements, standard real-time B-mode imaging was used for probe placement over inclusion, and then it was fixed over the target using a handle. Detailed information about the phantom is provided in Appendix A1.

To evaluate the performance of the system, a number of quality metrics were calculated. The shear wave motion energy in each pixel $SW_E(x,z)$ was calculated as:

$$SW_E(x,z) = \sum_{i=1}^{N} v_{zf}(x,z,i)^2, \tag{4.1}$$

where N is the number of samples in the slow-time domain and v_{zf} is the shear wave axial velocity after the directional filtering. The $SW_E(x,z)$ maps were calculated separately

for both the rightwards and leftwards propagating waves datasets and were then added together.

In the homogeneous phantom experiment, a bias and an SNR of the stiffness map were obtained as follows:

$$Bias = \frac{SWS_E - SWS_N}{SWS_N} \cdot 100\%, \tag{4.2}$$

$$SNR_{dB} = 20 \log_{10}\left(\frac{\mu_{SWS}}{\sigma_{SWS}}\right), \tag{4.3}$$

where SWS_E and SWS_N are the estimated and nominal shear wave speed values, respectively, μ_{SWS} is the mean value, and σ_{SWS} is the standard deviation of the selected area of the stiffness map.

Figure 4.5: The setup that was used in this work. Left to right: the L7-4 probe over the CIRS 049A elastography phantom, the us4R-lite portable research platform with the TX extension board mounted and connected to the PC.

In the heterogeneous phantom experiments, contrast and contrast to noise ratio (CNR) was derived using the mean SWS values of the inclusion and the background, μ_I and μ_B, and, in the case of (CNR), the standard deviations of those values, σ_I, and σ_B, within manually extracted regions of the stiffness map:

$$\text{Contrast} = \frac{|\mu_I - \mu_B|}{\mu_B}, \tag{4.4}$$

$$CNR_{dB} = 20 \log_{10}\left(\frac{|\mu_I - \mu_B|}{\sqrt{\sigma_I^2 + \sigma_B^2}}\right). \tag{4.5}$$

4.2.2   Results

This section is organized as follows. The first part shows the results of data processing at key stages of the reconstruction algorithm which is followed by the results of homogeneous and heterogeneous phantom experiments, respectively.

4.2.2.1   B-mode images

Figure 4.6 shows reconstructed B-mode images of various elastography phantom regions obtained as the compound of a set of angled plane wave acquisitions. Although full reconstruction of the B-mode images is not required for SWS map reconstruction, an inspection of the B-mode images was done to ensure proper probe placement, as discussed earlier. One can see, that inclusions of stiffness that differed from the background were all visible in B-mode images. In all cases, the contrast was poor and was almost the same for all inclusion types and diameters. Although B-mode imaging did not give any information that allowed to distinguish inclusion types, the inclusion shape and diameters were reconstructed very well.
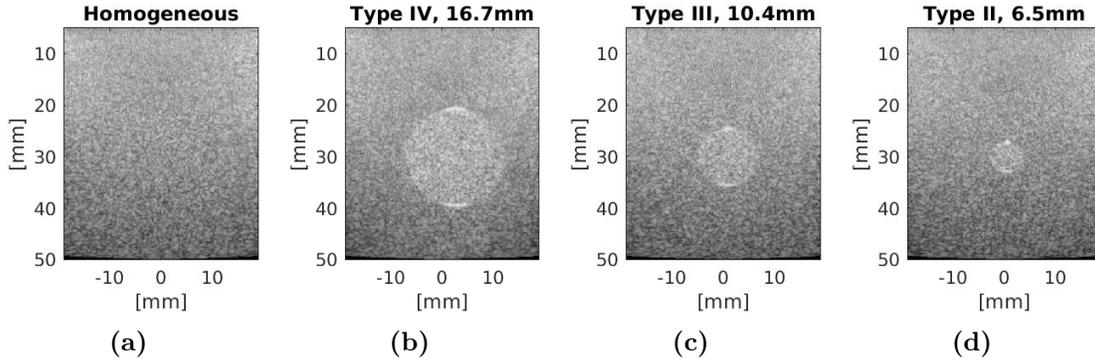
Figure 4.6: B-mode images of various phantom regions. (a) A phantom region with no inclusion (background only) used in the homogeneous experiment. (b) Type IV inclusion: 70.9 kPa nominal Young's modulus, 16.7 mm diameter. (c) Type III inclusion (34.0 kPa, 10.4mm). (d) Type II inclusion (8.6 kPa, 6.5mm).

4.2.2.2   Shear waves detection

Fig. 4.7 presents the shear wave motions detected in a homogeneous region of the phantom from each of the SSI acquisitions. Although the algorithm reconstructed the shear wave motion with a frame period of $0.2\,\mathrm{ms}$, only selected frames are presented to better show the wavefronts movement. The successive push beams induced clearly visible, broad shear wavefronts that covered most of the FOV. The peak particle axial velocity obtained was near $3\,\mathrm{mm/s}$. The time required for a wavefront to traverse the entire FOV was approximately $10\,\mathrm{ms}$. It can be seen that the shape of the wavefronts remained unchanged, however, their magnitude decreased along the propagation path, which was caused by attenuation. This effect justifies the usage of push beams generated at 3 lateral positions in the SSI method — it allows to suppress the shear wave attenuation and helps to cover the big part of FOV with shear waves of high magnitude that can be successfully detected. One can spot, that shear waves are not present on the pushing beam axis. The width of this region is a result of the time required for the generation of

push pulses and the time required for the scanner to switch into tracking mode. That also explains why the wavefronts are angled — when the second and third push pulses were generated, the shear waves induced by the first beam were already propagating.
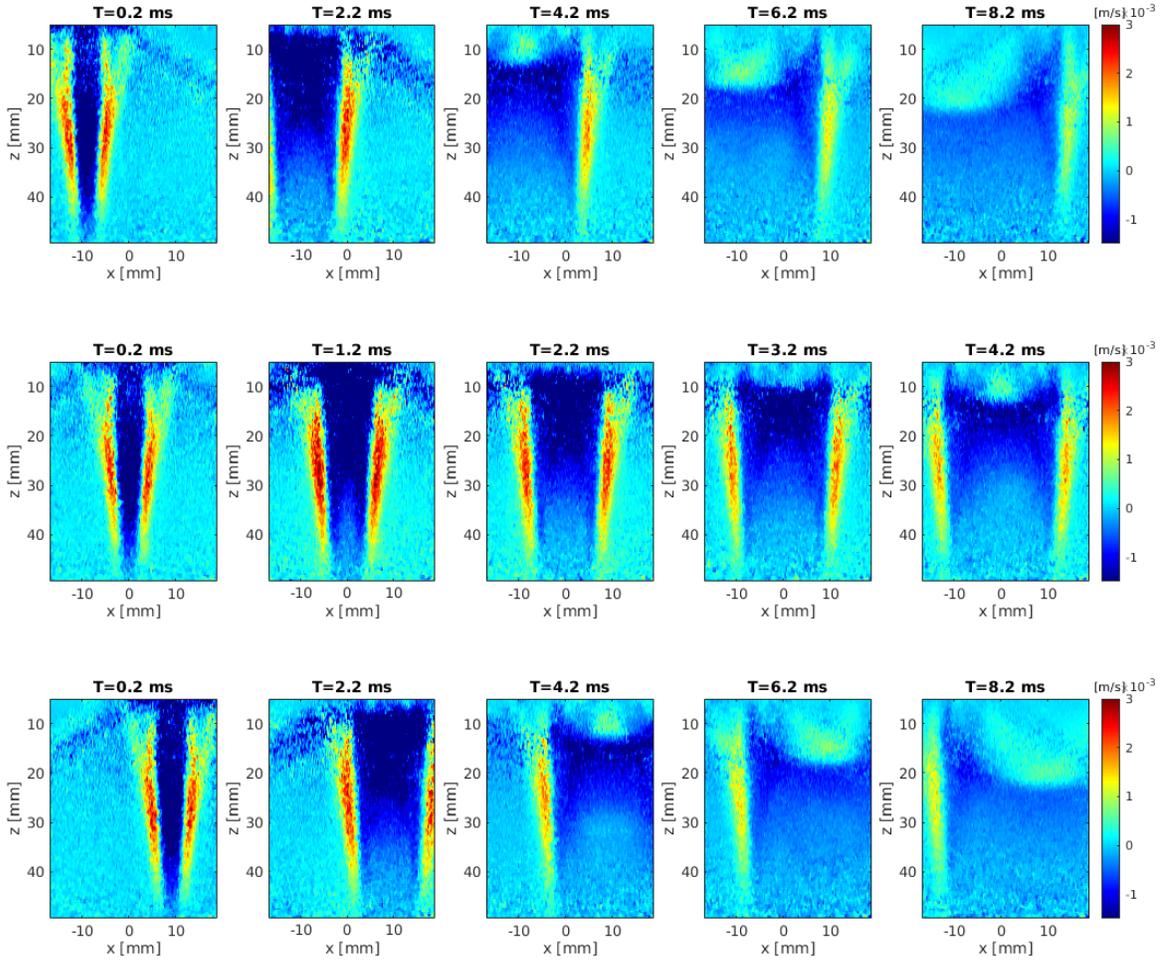
Figure 4.7: Shear wave motion data captured in the homogeneous region of the phantom. Each frame shows a transient particle axial velocity map. Top row: SSI sequence 1, middle row SSI: sequence 2, and bottom row: SSI sequence 3. The first image in each row shows the first frames captured in a given acquisition. The frame period in this image is 1 ms (every fifth frame is drawn) for the middle row and 2 ms for others (every tenth frame is drawn).

Due to longitudinal wave absorption of both push pulse and tracking pulses along their propagation paths, the greater was the depth, the shear wave detection noise was more apparent. This effect was especially noticeable at depths below 40 mm, where the shear wavefront was barely visible, despite the fact, that the last pushing beam was focused at the depth of 40 mm. In order to further expose this effect, additional acquisitions were performed to capture shear wave motion induced by individual beams of the SSI pushing sequence, focused at successive depths. The results are shown in Fig. 4.8. The shear wavefront induced by a first push was spread axially over a range of about 15 mm. In the case of second and third push pulses, the highest shear wave magnitude was detected close to the focal points, however, the shear waves were spread

over a big range axially. Even though the second push had a deeper focus than the first one, the shear wave magnitude near the focus was similar to the first push, possibly due to the fact, that the probe elevation focus was at 25 mm. This experiment shows the difficulty of delivering the energy to induce shear waves at greater depths.
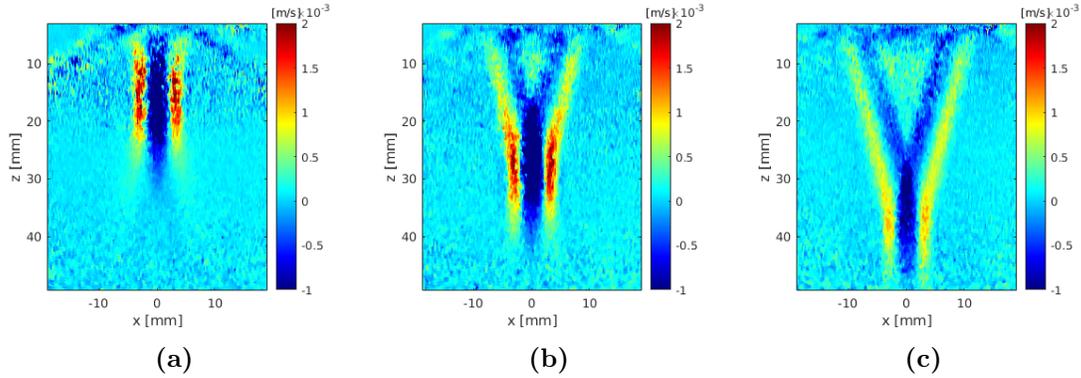
Figure 4.8: First frames of shear wave motion data induced by individual beams of the SSI pushing sequence. All the pushing beams used the same push length and the same f-number. (a): Push focused at 15 mm; (b): 30 mm; (c) 40 mm.

As discussed earlier in section 2.3.3, a selection of cut-off frequency for low-pass shear wave motion data filtering should be preceded by spectral analysis. Averaged spectral content of axial velocity profiles was obtained (taken from acquisition shown in Fig. 4.7, top row) in manually selected 1 mm x 1 mm ROI are shown in Fig. 4.9. The shear bandwidth was found to be nearly 150 Hz. It should be noted, however, that the spectral content of the shear waves may be dependent on ROI selection within a single acquisition. Since, in general, it is also dependent on push generation parameters and medium characteristics, a 500 Hz cutoff frequency picked for low-pass filtering in the reconstruction algorithm was high enough to not filter out the useful signal components.
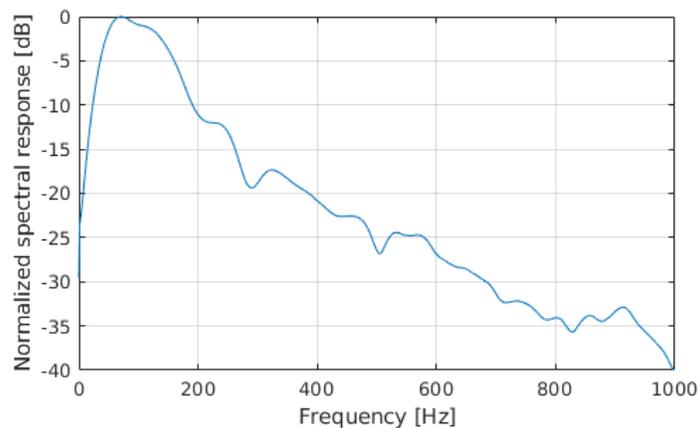
Figure 4.9: Normalized power spectra of shear waves averaged over 1 mm x 1 mm ROI 15 mm away from the push beam, at depth of 25 mm. The -6dB corner frequencies were 33 Hz and 167 Hz.

The results of shear wave motion data directional filtering in $[k - \omega]$ space are shown in Fig. 4.10 and Fig. 4.11. Except for the separation of shear waves propagating in

opposite directions, directional filtering improved the shear wave motion data SNR, which is apparent as improved contrast in Fig. 4.10 and Fig. 4.11.

(a)   (b)   (c)
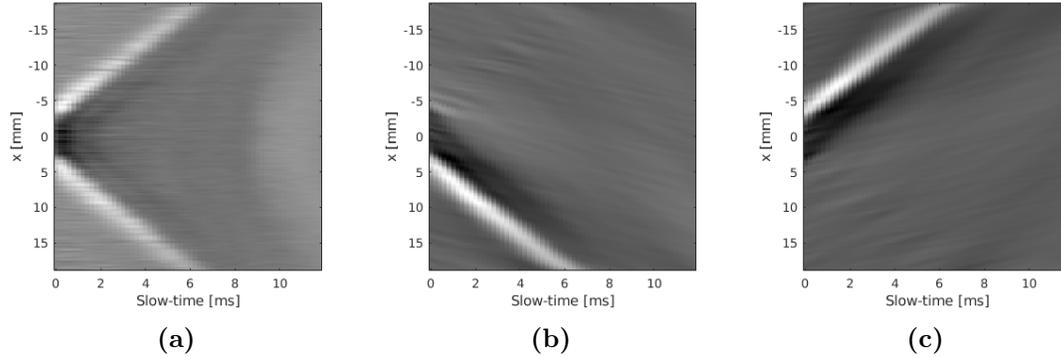
Figure 4.10: Example input and output data for directional filtering. (a): Input data — a shear wave motion data slice across the axial direction. The presented slice is taken from the acquisition shown in Fig. 4.7, middle row, and depth of 25 mm. A single shear wave source is visible at $x = 0$ and two waves are propagating toward the edges of FOV. (b) and (c): Directional filtering output data slices. Leftwards and rightwards waves are filtered out on (b) and (c), respectively. The same operation is performed for slices from each depth. The result of this operation is presented in Fig. 4.11.

After directional filtering, the next algorithm step involves stiffness map reconstruction using a correlation-based time-of-flight (ToF) estimation algorithm. In this step, the shear wave axial velocity profiles in slow time are analyzed. Fig. 4.12 presents example profiles from 3 locations along the lateral direction. There are two sets of profiles presented - one from the depth of 25 mm, and a second one from the depth of 45 mm. One can observe in Fig. 4.12(a) how the amplitude of the motion profile decreases along the propagation path due to attenuation. However, the shape of the profiles is preserved, which leads to high correlation coefficients in correlation-based ToF estimation. Peaks of the profiles can be also easily identified. As shown in Fig. 4.12(b), in the case of profiles from greater depths, the amplitude of particle velocity is comparable to noise, even despite filtering, and the ToF estimation using either correlation of the time-to-peak (TTP) method can lead to significant inaccuracies.

In order to further investigate the effect of shear wave attenuation, the energy distribution of shear waves generated by a complete SSI pushing sequence was obtained using Eq. (4.1). The shear wave energy distributions resulting from each of the 3 pushing sequences (in 3 various lateral locations) are shown in Fig. 4.13. The maps reveal how the shear waves were attenuated along their propagation paths and also show that the energy delivered to depths greater than 40 mm was very limited. Although a sequence of 3 laterally spaced beams complemented each other by covering the regions of axes of adjacent beams, it is also evident in Fig. 4.13(a) and Fig.4.13(c), that shear waves of significant energy are propagating outside the FOV. Since the SWS estimation uncertainty may depend on local shear wave energy, as stronger displacement signals may correlate better, it should be also noted, that during stiffness map reconstruction the data captured in each of these acquisitions were used to compound a single image described earlier.
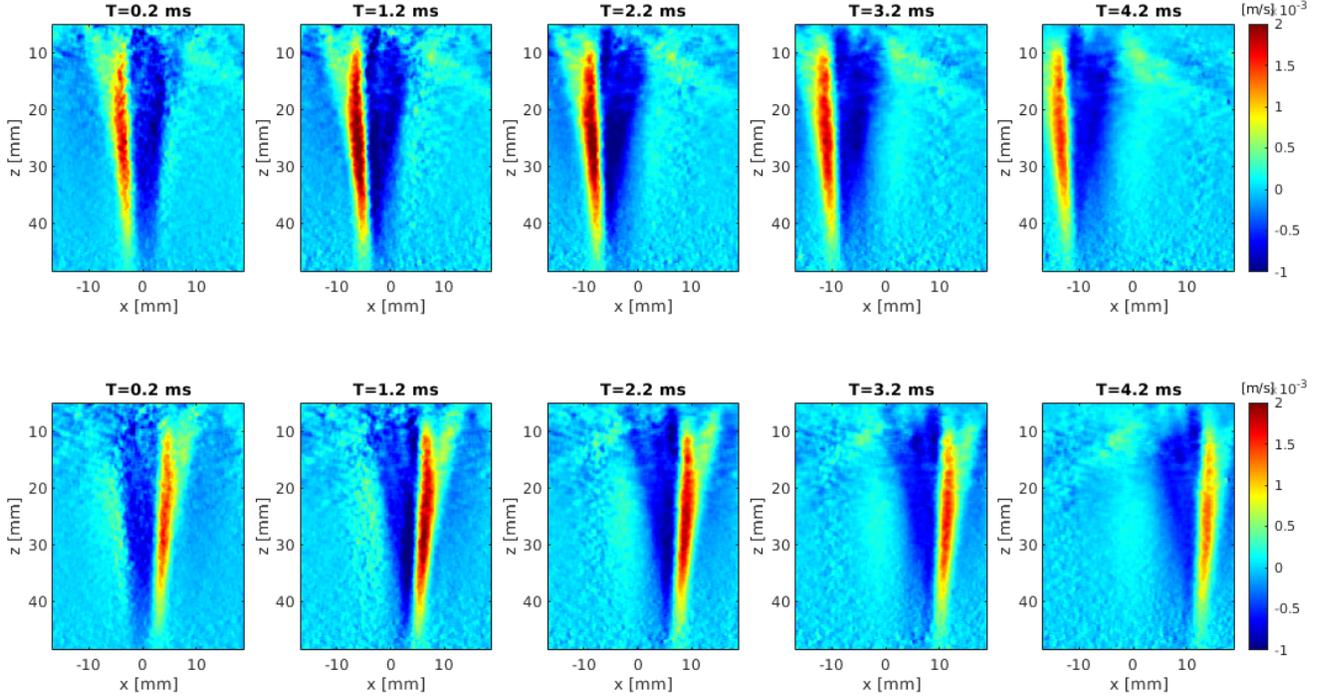
Figure 4.11: Shear wave motion data frames after directional filtering. The input motion data was taken from the acquisition as shown in Fig. 4.7, middle row. Top row: a result of extraction of leftwards propagating waves. Bottom row: rightward propagating waves extracted.

4.2.2.3   Homogeneous phantom experiment

In the homogeneous phantom experiment, the stiffness map acquisition was repeated in 5 different locations within the homogeneous region of the phantom. Fig. 4.14 shows the results of the experiment. A large ROI was manually selected to assess the quality metrics of the image. Results listed in Table 4.2 show good repeatability of the estimates — the difference in estimated average SWS within a ROI between any two out of 5 consecutive experiments was smaller than $0.03\,\text{m/s}$. However, considering the nominal value of homogeneous phantom SWS of $2.36\,\text{m/s}$ (as declared by the manufacturer, with $\pm 5\%$ tolerance of nominal Young's modulus), in all cases average SWS estimates were biased high. SNR values derived in ROIs were in the range of 29.7 to 32.1 dB. One can also observe that the measurement exhibited a significant amount of noise at depths below $40\,\text{mm}$. As discussed in the previous section, it was caused by insufficient shear wave energy at this depth. For this reason, that region was excluded from the selected ROI.

Fig. 4.14(b) shows the pixel-wise distribution of bias. It was obtained as the difference between estimated SWS and nominal background SWS value declared by the phantom manufacturer. A display range was adjusted to expose the bias pattern. The image obtained this way revealed, that regions exhibiting the highest bias overlap the pushing beams' axes. Fig. 4.14(c) presents the distribution of the SWS estimates within the ROI. Estimated SWS values were within the range of 2.2 to $2.8\,\text{m/s}$ with a mean of $2.47\,\text{m/s}$ and a standard deviation of $0.08\,\text{m/s}$.
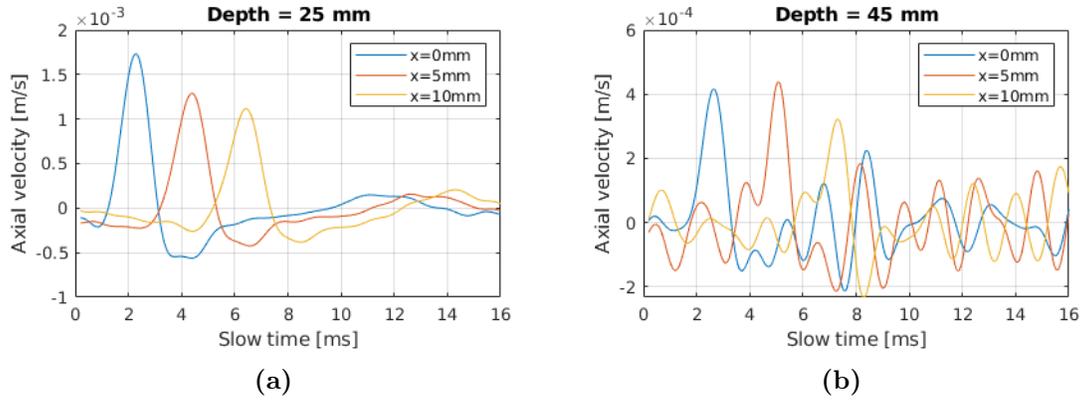
Figure 4.12: Shear wave axial velocity profiles vs. slow time in three lateral locations and from two depths. The profiles are taken from motion data as shown in Fig. 4.11, bottom row. (a): Profiles from the depth of 25 mm. (b) Profiles from the depth of 45 mm.
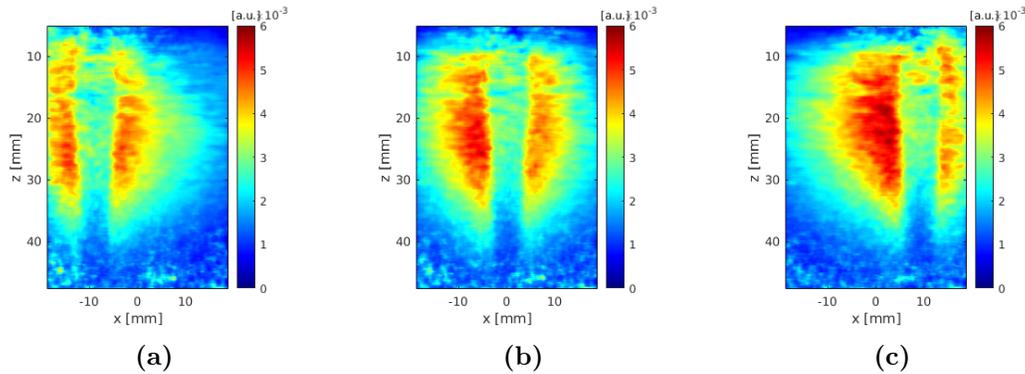
Figure 4.13: The shear wave energy distributions resulting from each of 3 pushing sequences (in 3 various lateral locations) that build together the excitation used in the SSI method. The square root of total energy obtained using Eq. (4.1) is shown.
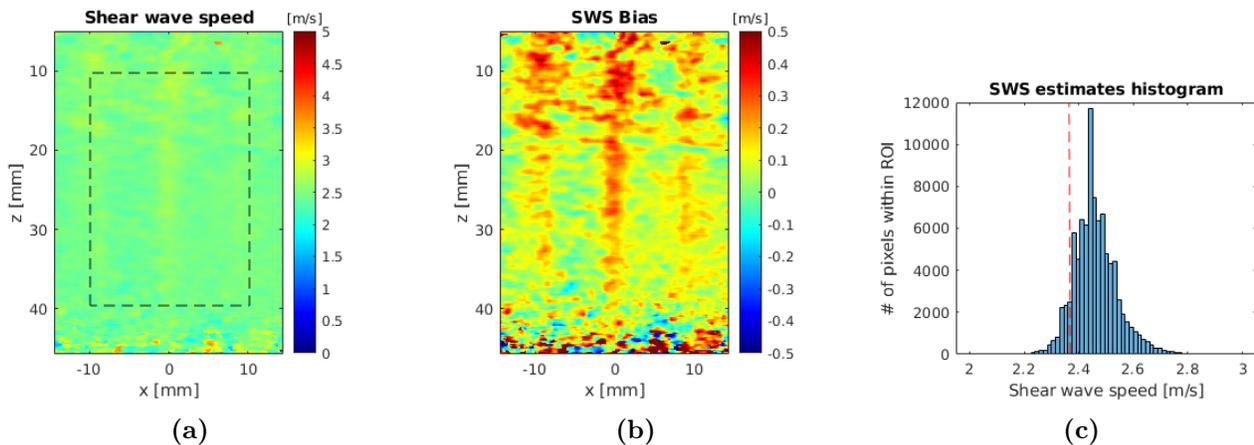
Figure 4.14: Homogeneous phantom experiment results. (a): An example SWS map obtained in one of 5 acquisitions in the homogeneous phantom experiment. ROI used to obtain image quality metrics is marked in the figure with a dashed line. Detailed results can be found in Table 4.2. No median smoothing was applied on this map. (b): SWS bias of the map in (a). (c): Histogram of SWS estimates obtained within the ROI. The dashed red line shows the nominal value of SWS (declared by the phantom manufacturer).

Table 4.2: Homogeneous experiments results summary.

Parameter	T1	T2	T3	T4	T5	Avg.
Average SWS [m/s]	2.47±0.08	2.48±0.08	2.47±0.08	2.47±0.08	2.45±0.06	2.47±0.13
Average bias [m/s]	0.12 (+4.9%)	0.12 (+4.9%)	0.11 (+4.6%)	0.11 (+4.6%)	0.10 (+4.2%)	0.11 (+4.6%)
SNR [dB]	29.7	30.3	29.9	30.4	29.9	32.1

While the histogram in Fig. 4.14(c) gives information about the variance of the estimates from pixels within the ROI in an image coming from a single acquisition, Fig. 4.15 presents the SWS estimate variance maps obtained from several acquisitions to visualize how the estimation precision depended on pixel position. There were two cases investigated. In the first one (Fig. 4.15(a)), the probe was placed over ten different regions of the homogeneous phantom to capture SWS maps corresponding to various speckle distributions. In the second case (Fig. 4.15(b)), the probe position was fixed over the same phantom region and the acquisition was repeated ten times. As expected, the highest results deviations were found in regions of limited shear wave energy. Fig. 4.15(c) and Fig. 4.15(d) illustrate the distribution of standard deviations of the estimates within the same ROI as used so far for two investigated cases. The average standard deviation in a ROI of variable probe position case was 0.044 m/s, whereas in the case of fixed probe position, it was more than twice smaller — 0.02 m/s. Interestingly, in the latter case, almost 30% of pixels within the ROI had standard deviations of SWS estimates very close to or equal to 0.

4.2.2.4   Heterogeneous phantom experiment

In the heterogeneous phantom experiment, the ability to detect inclusions of stiffness both higher and smaller than the background was evaluated. The results of the experiment are presented in Fig. 4.16. No spatial filtering was applied to these maps. The detailed results are listed in Table 4.3. Visualization of the inclusion SWS estimates is shown in the form of bar plots in Fig. 4.17.

SWS maps obtained for Type IV inclusion show good contrast, especially for the biggest diameter of inclusion. The SWS estimate of the inclusion decreased along with the diameter. For example, inclusion type IV SWS estimate dropped from 4.53 m/s to 3.73 m/s from inclusion diameter 16.7 to 6.5 mm, so by nearly 18%. All results are biased low - minimum by 7.5% up for inclusion size 16.7mm to 23.2% for the smaller diameter. Due to the smaller nominal SWS difference between inclusion and background material, the inclusions type III SWS maps showed worse contrast than in the previous case. The same trend could be observed - the inclusion SWS estimate decreased along with the inclusion diameter. This effect in case of the smallest inclusion size made the inclusion barely detectable in the image, as the artifacts in the background featured comparable sizes and signal levels. In the case of Type II inclusions SWS maps, all the inclusions could easily be seen, since the inclusion material's low SWS value did not appear in any other image region. In contrast to the previous two cases, type II inclusions images exhibited reversed trend of inclusion diameter effect on SWS estimate. Here, estimated inclusion SWS increased along with decreasing inclusion diameter. In addition, all the results were biased high. The same applied to Type I inclusions SWS maps. In the
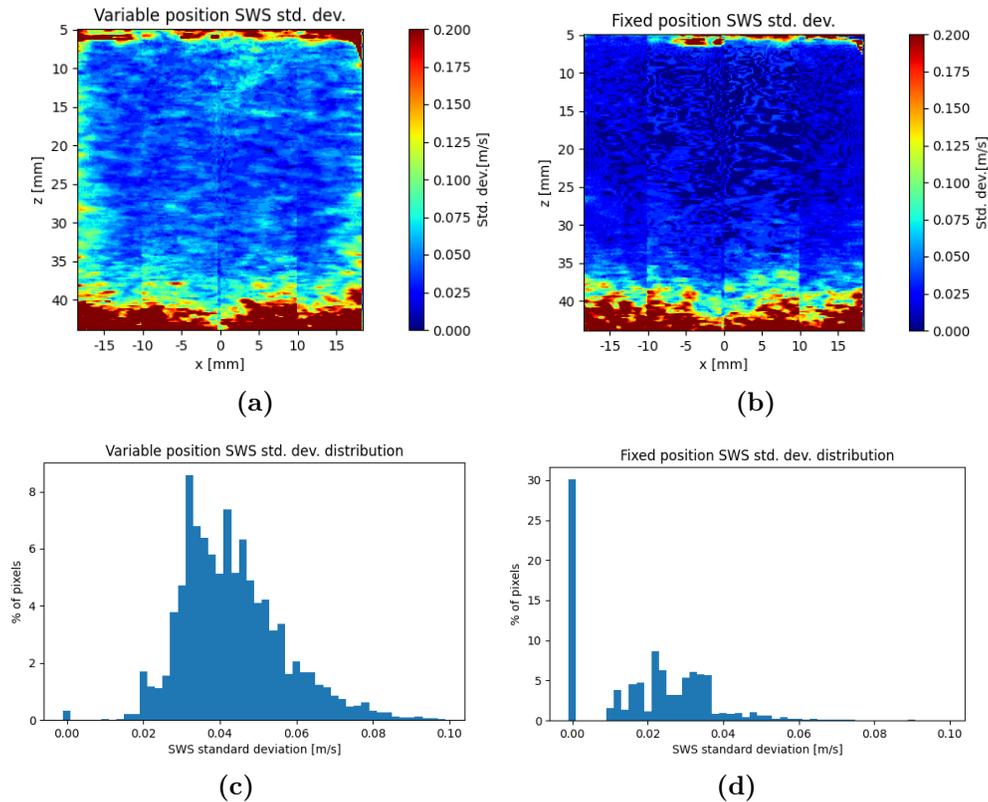
Figure 4.15: SWS estimation precision evaluation in the homogeneous experiment. (a): Variance map of the SWS estimates obtained from 10 acquisitions done at different phantom locations. (b): Variance map of the SWS estimates obtained from 10 acquisitions done at the same phantom location. In each map, each pixel shows the variance of corresponding pixels SWS estimates from 10 SWS maps. The same colormap was used in images (a) and (b). (c): Histogram of values of the image (a) within the ROI as used in homogeneous experiment quality metrics assessment (see 4.14(a)). (d): Same as in (c), but obtained for image (b).
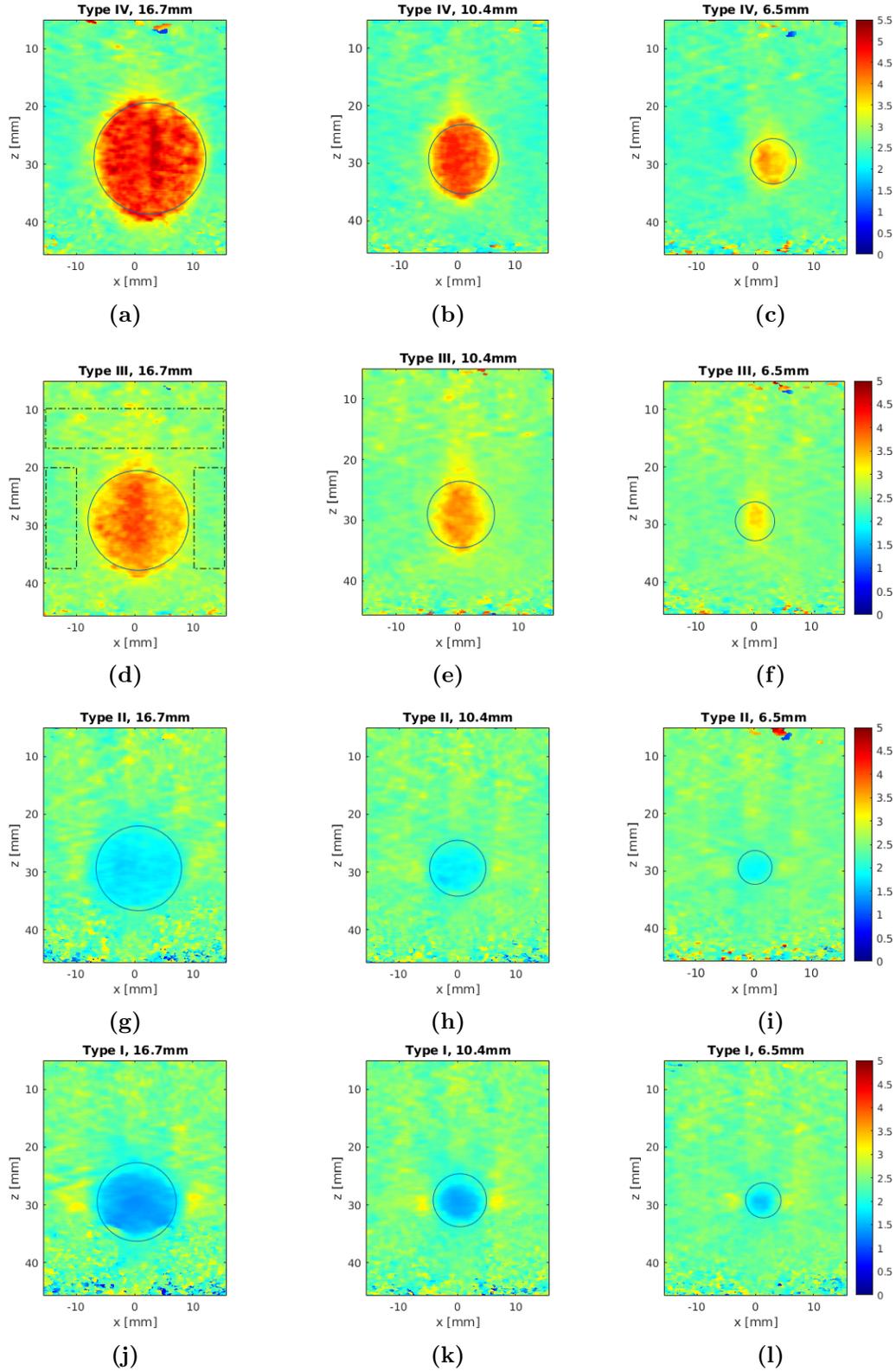
Figure 4.16: Heterogeneous phantom experiment results — stiffness maps of inclusions of various stiffness and diameters. Top row — inclusion type IV, second row — inclusion type III, third row - inclusion type II, bottom row — inclusion type I. Black circles show positions of inclusions obtained by B-mode imaging. The color map shown in the rightmost images is valid for all images within the same row. Detailed quality metrics of the presented maps can be found in Table 4.3.

105

case of type I inclusions SWS maps one can spot that stiffness maps underestimate the diameter of inclusion. For bias expressed in percentages, type I inclusion average SWS estimates suffered the highest levels of bias (up to 32%), however, the highest absolute error 1.13 m/s was for type IV inclusion of 6.5 mm diameter.

The background SWS estimate was biased high in all measurements, with bias ranging from 3.6% up to 8.4%. Despite the biggest diameters of inclusions, this bias was comparable to that measured in the homogeneous phantom experiment.
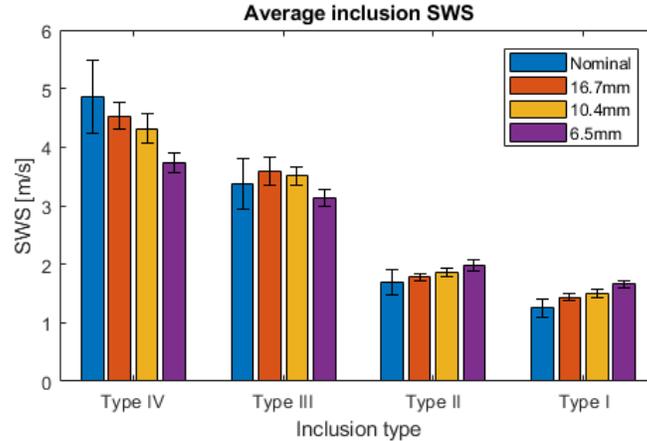
Figure 4.17: Comparison of heterogeneous phantom experiment results for various inclusion stiffnesses and diameters. Standard deviations of SWS estimates within ROIs are used as error bars. In the case of nominal values, tolerance of inclusion stiffness as specified by the manufacturer was used to draw the error bars.

4.2.3 Discussion and comparison with other systems

A) On system validation

The system validation experiments results confirm, that the proposed system is capable of implementing the 2-D SWE technique. It was achieved by satisfying two key requirements — generation of high-energy push pulses and high frame-rate data acquisition of shear wave propagation. Hardware capability of pushing beam generation was proven using a well-established SSI method, that features a kind of complexity — it requires a generation of a sequence of beams with strict timing. The presented system's transmit flexibility turned out to be good enough to implement this method. The second requirement for the 2-D SWE was achieved by PWI data capture with a 5 kHz frame rate. Data quality allowed small axial displacement detection of shear waves with various energy levels. It enabled full SWS map reconstruction using the correlation-based ToF SWS estimation using the developed algorithm, preceded by 3-D shear wave motion data filtering in both time and Fourier domains to improve conditions for correlation calculation.

Moreover, it was shown, that the developed system can be used to obtain data at various stages of the processing and to conduct various experiments aimed at evaluation of various 2-D SWE methods. Although a single imaging scenario was presented here for clarity, similar experiments can be performed for other push generation methods or data acquisition schemes. An example of such research was presented in the article published

Table 4.3: Heterogeneous experiments results summary

Parameter	16.7mm	10.4mm	6.7mm
Inclusion type IV (nominal SWS = 4.86 m/s)			
Inclusion SWS [m/s]	4.53±0.28	4.32±0.25	3.73±0.16
Inclusion bias [%]	-7.5	-11.1	-23.2
Inclusion SNR [dB]	24.0	24.8	27.0
Background SWS [m/s]	2.51 ±0.13	2.42 ±0.12	2.42±0.08
Background bias [%]	+6.7	+4.4	+3.6
Background SNR [dB]	25.7	26.2	29.8
Contrast	0.81	0.79	0.53
CNR [dB]	16.2	16.8	17.0
Inclusion type III (nominal SWS = 3.37 m/s)			
Inclusion SWS [m/s]	3.60±0.24	3.51±0.16	3.13±0.14
Inclusion bias [%]	+7.9	+5.5	-6.9
Inclusion SNR [dB]	23.6	27.2	26.8
Background SWS [m/s]	2.46 ±0.13	2.42±0.09	2.42±0.12
Background bias [%]	+5.6	+3.7	+4.4
Background SNR [dB]	25.6	28.2	28.3
Contrast	0.46	0.45	0.30
CNR [dB]	12.5	15.7	12.4
Inclusion type II (nominal SWS = 1.69 m/s)			
Inclusion SWS [m/s]	1.78±0.06	1.86±0.07	1.98±0.10
Inclusion bias [%]	+6.2	-9.7	+16.9
Inclusion SNR [dB]	29.5	20.90	18.27
Background SWS [m/s]	2.43 ±0.18	2.43±0.14	2.43±0.12
Background bias [%]	+5.2	+5.2	+4.5
Background SNR [dB]	22.6	25.1	26.4
Contrast	0.26	0.24	0.18
CNR [dB]	10.3	11.3	9.2
Inclusion type I (nominal SWS = 1.25 m/s)			
Inclusion SWS [m/s]	1.44±0.07	1.51±0.07	1.66±0.06
Inclusion bias [%]	+15.3	+21.1	+32.4
Inclusion SNR [dB]	26.0	26.8	28.7
Background SWS [m/s]	2.47 ±0.22	2.43±0.14	2.39±0.11
Background bias [%]	+8.4	+5.6	+3.8
Background SNR [dB]	20.9	24.6	26.5
Contrast	0.42	0.38	0.31
CNR [dB]	12.9	14.5	15.1

by the author of this thesis [80].

B) Bias of the SWS estimates

In this work, the SWS values declared by the phantom manufacturer were treated as nominal values. Those values had a 5% tolerance for a given stiffness. Nevertheless, in the homogeneous experiment, all the estimated average SWS values obtained in all tries ranged 2.45–2.48 m/s and were beyond this margin, which is 2.41 m/s maximum for background material. The bias with respect to the nominal value was between 4.2–4.9%. Considering the possible bias sources as described in section 2.5.2, the bias distribution pattern shown in Fig. 4.14(b) with the bias dependent on depth and lateral positions suggests, that the bias is a result of the pushing beam shape or diffraction. Zhao et al. [156] used the Verasonics research scanner and performed bias measurements using the same probe as in this thesis (ATL/Philips L7-4) and various focal configurations and 3 homogeneous phantoms of varying nominal stiffness. They reported bias of similar levels (up to ∼8%). Bias was reported to vary with respect to depth and focal configuration. In this thesis, even higher levels of bias were found in heterogeneous phantom results in most cases, ranging between -23.2 to 32.4%. A rough estimation of shear wavelength based on spectral analysis shown in Fig. 4.9 and assumption of nominal background SWS leads to a conclusion, that it is in the range of over a dozen millimeters. This explains SWS underestimation of inclusions stiffer than the background and SWS overestimation of inclusions softer than the background. Similar results were reported by Racedo et al., who used Verasonics scanner, L7-4 probe, and CUSE method for imaging of heterogeneous inclusions of the same elastography phantom [140]. They reported bias of levels between -40 and 5%, depending on inclusion stiffness, size, and reconstruction parameters. An even more extensive study was held by Kishimoto et al., who compared the results obtained from homogeneous viscoelastic phantom by 6 different SWE-capable ultrasound systems and various probes [196]. For example, at the mean value of 2.23 m/s, the results from different systems ranged between 1.99 to 2.64 m/s at depth of 2 cm for a linear array. Other works show that the the estimation result in heterogeneous cases may be operator-dependent [197], and target size, depth and stiffness-dependent [198]. Another interesting article compared Siemens Acuson S2000 and SuperSonic Imagine Aixplorer systems for *in-vivo* breast imaging of same targets [199]. They reported differences reaching nearly 50% between these machines. A similar study was held by Song et al. [200], who compared Aixplorer system with General Electric LOGIQ E9 in phantom with inclusions imaging and reported biases of both machines, in range of -28 to 38% for LOGIQ E9 and -24 to 36% for Aixplorer, while bias levels were again dependent on inclusion stiffness and size. To summarize, identifying the bias sources and the bias levels they introduce may require an additional, detailed, and systematic analysis.

C) Precision of the SWS estimates

The same applies to estimates variance. As shown in section 2.5.1, there are many parameters affecting the SWS estimation accuracy. The obtained results show that the standard deviation of SWS estimates in the homogeneous region was in range of 0.06 to 0.13 m/s when analyzing estimates distribution within a ROI of a single image. Repeated acquisitions experiment showed that single pixels have average estimate deviations of 0.02 m/s and 0.044 m/s in fixed and variable probe position cases, respectively. This shows how significant part of estimation uncertainty is brought by the speckle bias effect.

In the case of inclusions imaging, SWS estimates standard deviation increased along with inclusion stiffness from 0.07 to 0.28 m/s in inclusion types I and IV, respectively, which in general is in line with the uncertainty model of Deffieux.

D) On the resolution of the SWS maps

There was no systematic evaluation of SWS maps resolution held in this validation study. As described in section 2.5.3, two key factors that define the 2-D SWE method resolution are shear wavelength and lateral kernel size used in ToF SWS estimation. Regarding shear wavelength, assuming a wave propagation speed of 2.35 m/s (nominal for phantom background material) and shear wave center frequency of 100 Hz (see Fig. 4.9), the wavelength can be estimated to be around 2.35 cm. Importantly, however, there is a trade-off between accuracy and spatial resolution that is adjusted by the latter factor. In this study, only a single kernel size was used — 3.2 mm in the heterogeneous experiment that was found experimentally as a reasonable compromise. The non-zero size of lateral estimation kernel size also averaged the estimation results within the kernel range. This effect was apparent in reconstructed SWS maps as blurring of inclusion edges, which degraded the image resolution. Despite that, SWS maps of various inclusion types and diameters (Fig. 4.16) give an overview of the system's capabilities in terms of resolution. For instance, a 6.5 mm inclusion type III can barely be seen in the estimated SWS map due to poor contrast, whereas inclusions type IV or type II of that size were easily detectable. To expect better resolution, one may try to use different push generation sequence with shorter push lengths and use smaller lateral SWS estimation kernel sizes.

4.3 Processing performance evaluation

One of the aims of this thesis was to develop a system solution of the performance providing real-time 2-D SWE imaging capabilities. The system architecture and the implementation process targeted to fulfill this aim were described in section 3.4.5. In this section, the developed system performance in terms of image reconstruction speed is evaluated. The factor of power consumption of the system is also addressed.

4.3.1 Methods

In general, during active operation of the imaging system, the total time required to obtain a single image in the SWE method consists of two components — time required to acquire data of the imaging sequence and time required to reconstruct an image from raw RF data. The former component involves push generation and shear waves tracking sequence in a procedure that is application-specific and it typically ranges between tens of milliseconds up to a few hundred milliseconds, depending on the number of acquisitions required to reconstruct a single image, medium characteristics, or the size of ROI. Although it is not related to the reconstruction performance, it should be noted that acquisition sequence parameters define the amount of data for processing, dictating this way the size of the computational problem. The latter component — reconstruction time — is the main target for the evaluation in this section, since, in contrast to the former component, reconstruction time can be optimized by adjusting reconstruction parameters.

A)   Evaluated implementation types

Throughout the software development process of the reconstruction algorithm, three different fully-featured implementations were created. These can be briefly characterized as follows:

- Implementation in Matlab environment (The Mathworks, Natick, MA, USA). It was created in the early stage of development to verify the first algorithm prototype and to experimentally adjust reconstruction parameters. Except for vectorization, no performance optimization techniques were used in this implementation. This version will be referred as "*Matlab implementation*".

- As the first step toward reaching real-time performance, the developed algorithm was transferred to Python language. One exception was beamforming step, which was not coded in Python, since the GPU version was already available. This version used *NumPy* and *SciPy* libraries for numerical operations. Since this version was only a middle step before migrating the algorithm to GPU-based processing, careful optimization of the code was not performed, although it should be noted, that *NumPy* and *SciPy* libraries use multi-threading of some operations by default. This implementation will be referred to as "*CPU implementation*".

- The last version used GPU to enhance processing performance. It was developed in Python and involved using *CuPy* library and custom GPU kernels to utilize GPU resources. Refer to section 3.4.5 for implementation details. This implementation will be referred to as "*GPU implementation*".

While the focus was put on GPU implementation, the two other implementations were also evaluated to explore the benefit of using GPU in this specific application. Unless otherwise mentioned, the default values of the algorithm were used (see section 3.3 for details).

B)   Hardware setup

As described earlier in Chapter 3, the MSI GS66 mobile workstation (Micro-Star Int'l Co. Ltd., Taipei, Taiwan) was used, which was equipped with Intel Core i9-10980HK CPU with 18 cores, Nvidia RTX3060 GPU (Nvidia, Santa Clara, CA, USA), and 32 GB of RAM memory.

C)   Experiments

A series of experiments were conducted to measure the processing time of the developed reconstruction algorithm. Unless otherwise mentioned, the input test dataset consisted of 100 RF data frames, each comprising 4224 samples from each of 128 receive channels captured at a sampling frequency of 65 MHz. The dataset was acquired using a pushing beam in the CUSE method, which consisted of three parallel beams, enabling the reconstruction of an elasticity map of the entire region under the probe (laterally) from a single dataset. For comparison purposes, all three aforementioned implementation versions were evaluated by measuring the total processing time required to obtain a single image. In each case, the same default values of reconstruction parameters were used for fair comparison (see section 3.3). To provide a comprehensive insight into data processing performance and to identify the most computationally intensive algorithm stages, the processing time of each stage was measured. In the second part, tests were

conducted to assess the impact of the size of the input dataset on the processing time of the whole GPU-accelerated algorithm and its individual stages. The size of the input dataset was controlled by both number of frames to be processed and by the granularity of the beamforming grid, which defined the size of the dataset for the SWE-specific reconstruction algorithm steps. Finally, in the last part, the algorithm stages with the highest computational complexity were identified, and tests were carried out for them to evaluate the impact of individual reconstruction parameters on processing time.

D) Measurements methodology

Depending on the implementation type, different methods were used to determine the time elapsed during data processing. For the Matlab implementation, well-known *tic/toc* functions were used. In CPU implementation the *timeit* package was used. In the case of GPU implementation a dedicated tool was used — Nvidia Nsight Systems performance analysis tool (Nvidia, Santa Clara, CA, USA). This profiler enabled tracing GPU computing tasks by reporting processing time by each launched kernel or providing insight into memory usage and throughput. In all three cases, the processing time was derived by running the algorithm code (or its part) in a loop with 10 iterations and measuring the time elapsed for each iteration. Average time along with standard deviation was reported for each test.

4.3.2 Results

A) Implementations performance comparison

Table 4.4 presents the comparison of performance achieved by three evaluated implementations. Most noteworthy, the GPU-accelerated implementation outperformed the remaining two by the order of around two magnitudes. By average, it required 348 ms to fully process the input dataset (including data transfers into and out of GPU), whereas CPU and Matlab implementations required 29.67 s and 36.11 s, respectively. This makes GPU implementation 104 times faster than Matlab implementation and a minimum of 85 times faster than CPU implementation since the latter did not feature a beamforming step.

Table 4.4: Performance comparison of Matlab, CPU, and GPU implementations. Default algorithm parameters were used. Average processing times out of 10 runs are shown.

Processing time [ms]				GPU speed-up	
Stage	Matlab	CPU	GPU	vs. Matlab	vs. CPU
Data transfers	0	0	14.20	N/A	N/A
RF Filtering	497.6	10753.3	36.5	14x	295x
Demodulation	1204.7	12220.0	54.5	22x	224x
Beamforming	30608.4	N/A	16.81	1821x	N/A
Angle compounding	88.5	44.8	0.5	134x	68x
Shear wave detection	851.7	168.9	3.5	246x	49x
Shear wave motion data filtering	406.4	480.7	23.22	18x	21x
SWS estimation	2438.7	5996.6	198.5	12x	30x
Post-processing	14.6	10.4	0.6	25x	18x
Total [s]	36.11	29.67	0.348	104x	85x

Measuring the processing time of individual algorithm's steps revealed, that in comparison to GPU implementation, the CPU version achieved much worse performance at early stages — RF data filtering and demodulation – being 295 and 224 times slower and spending most of the time (77.5%) in these steps. In the remaining, SWE-specific steps, the GPU implementation was found from 18 up to 68 times faster than in CPU-based alternatives. In contrast, the Matlab version performed quite well in the RF filtering and demodulation steps spending in these steps 1.7 s, which is a much better result than the CPU version, but still significantly worse than in GPU-accelerated application, which showed 18x speed-up in this parts combined. In the case of the Matlab version, beamforming was the algorithm step that contributed most of the total processing time (80%). In SWE-specific steps this version was found much slower in shear wave detection — 851 ms, almost 250x slower than GPU implementation — and over a dozen times slower in shear wave motion filtering and SWS estimation steps.

B)   Evalaution of GPU implementation performance
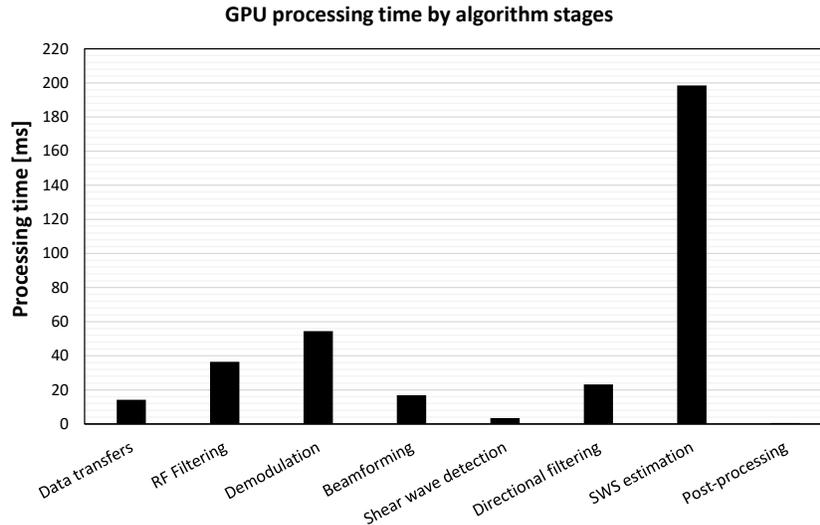
Figure 4.18: Contribution of each algorithm step to the total processing time of the GPU implementation. Default values for algorithm parameters are used. The data transfers bar includes both input and output data transfers.

The remaining part of this section will focus solely on the GPU implementation. The bar plot presented in Fig. 4.18 depicts how each of the algorithm's steps contributed to the overall processing time of the GPU implementation. The SWS estimation was clearly the most computationally intensive step that required almost 200 ms being this way responsible for more than half of total processing time (57% of total 348 ms). Within the remaining part, the biggest contributions were brought by demodulation (54.5 ms, or 15.6%), RF data filtering (36.5 ms or 10.5%), and shear wave motion data filtering (23.2 ms or 6.7%). Time brought by other stages, like beamforming, or shear wave detection and post-processing in particular was rather negligible. Please note, that the default algorithm parameters were used here — the evaluation of the impact of individual parameters on the processing time of selected algorithm steps is presented further in this section.

Fig. 4.19 shows how the total processing time is influenced by the computational problem size. The relation between processing time and input dataset number of frames is nearly linear (Fig. 4.19(a)) ranging from 166.8 ms for 50 frames to 578 ms for 150 frames. Assuming linear relation it gives around 4.1 ms per frame. Conversely, the relation between the beamforming grid step and the total processing time is not linear (Fig. 4.19(b)). However, the grid step parameter defines the number of target image pixels in both axial and lateral dimensions, which explains the shape of the curve. For example, grid steps of 0.2 mm and 0.4 mm resulted in a target image pixel number of 40 and 10 thousand, respectively. Obviously, the processing time increased when finer grids were used — 375 ms for the finest grid and 165 ms for the grid step of 0.4 mm. Worth noting, that the total processing time is not directly proportional to the number of pixels, since the algorithm stages that precede beamforming are not affected by the grid step setting.
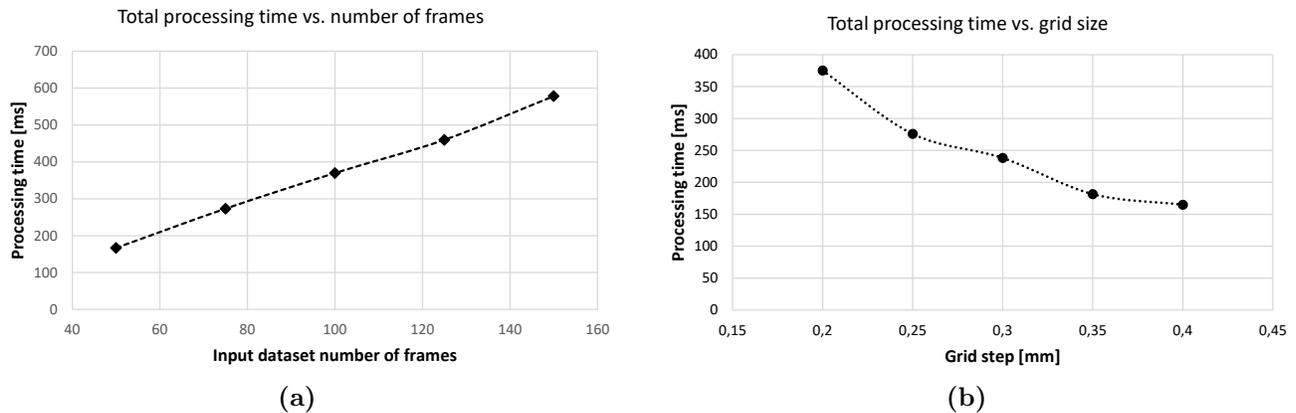
Figure 4.19: Evaluation of computational problem size impact on processing time. (a): Controlling input dataset size by defining a number of frames and keeping the grid step constant at 0.2 mm. (b): Controlling input dataset by beamforming grid step. The number of frames is kept constant at 100 frames.

The same relationship, but broken down into individual tasks is presented in Fig. 4.20. The processing time of each sub-task seems to be affected by the input dataset size in a similar way as the whole processing pipeline — it increases linearly along with the size of the input dataset. One exception is the directional filtering step. A linear relation is disrupted probably by the fact, that this step performs 2-D FFT and both dimensions are padded to the nearest power of two. One observation worth noticing is that even though SWS estimation remains the longest stage of the algorithm to perform, its relative contribution to the total processing time decreases when coarsening the beamforming grid. This is due to the fact, that the time contribution brought by stages preceding beamforming remains unchanged.

Although the input dataset dimensions influence greatly the computational workload and the total processing time consequently — which was shown above — there are also other reconstruction parameters that have a high impact on processing time. The results of measurements evaluating these effects are presented in Fig. 4.21. The plots shown in Fig. 4.21(a) and Fig. 4.21(b) show that the processing time of RF filtering and demodulation steps can be controlled by the FIR filters number of taps, and the relation between processing time and a number of taps is linear, what is rather an expected result. For instance, for an input dataset consisting of 150 frames reducing the number
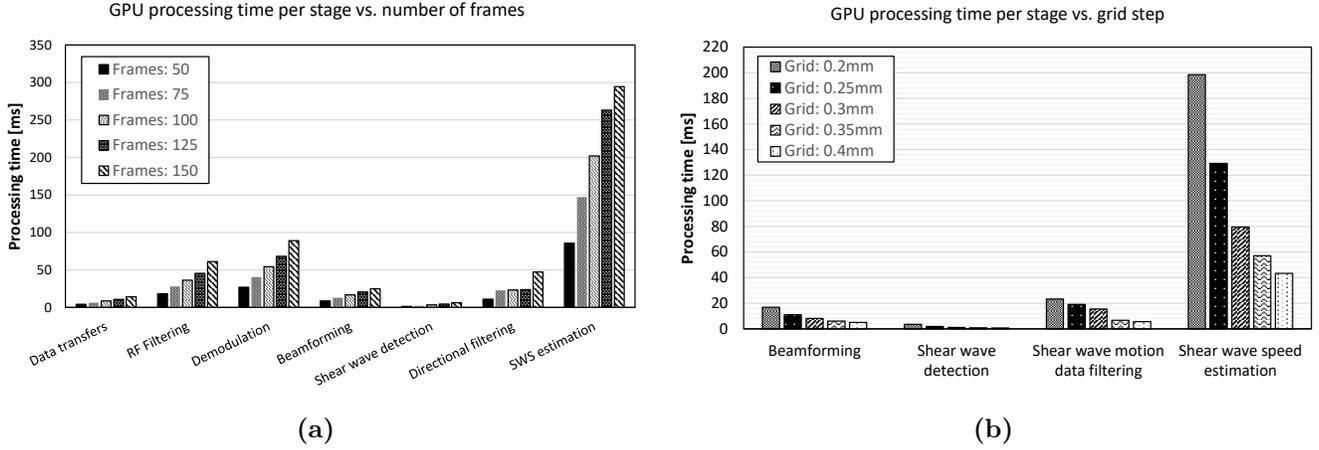
Figure 4.20: Evaluation of computational problem size impact on processing time broken down into algorithm stages. (a): Controlling input dataset size by defining a number of frames. Grid step is 0.2 mm (b): Controlling input dataset by beamforming grid step. Number of frames is 100. Only algorithm steps affected by grid resolution are shown here.

of taps from 256 down to 64 can reduce the processing time by 38 ms and 101 ms for RF filtering and demodulation, respectively. Characteristics plotted in Fig. 4.21(c) reveal that the Kasai algorithm is twice faster than Loupas' version of the shear wave detection method. In both cases, however, time spent in this stage is small in comparison to other algorithm steps.

On the other hand, the results shown in Fig. 4.21(d) highlight how significantly interpolation settings applied for the SWS estimation stage can affect the processing time. In general, the processing time increases linearly along with the interpolation factor in proportion to the interpolation order. For instance, the SWS estimation stage with 100 shear wave motion data frames requires 580 ms to execute when using order 3 spline interpolation with a factor of 10, whereas it needs only 200 ms to run with interpolation factor of 5 and spline interpolation order of 2, being almost 3 times faster.

4.3.3   Discussion

A)   On the achieved performance and real-time capabilities

The overall results demonstrated that the developed system solution exhibits a level of performance that enables real-time 2-D SWE imaging. However, it is important to note that this achievement was made possible through the utilization of parallel processing using a GPU. When comparing three different implementations (Matlab, CPU, and GPU) and after applying a typical set of reconstruction parameters, it became evident that only the GPU-based implementation can process the data within a reasonable timeframe, allowing for timely feedback. The GPU implementation outperformed the alternatives by completing the reconstruction under 350 ms (2.85 frames/s), while the others required over half a minute (Matlab) or nearly the same amount of time (CPU). The speed-ups achieved by utilizing GPU ranged between 18x up to nearly 300x versus the CPU implementation, depending on the algorithm stage, which clearly shows the benefit of GPU usage in this application.

(a)

(b)

(c)

(d)
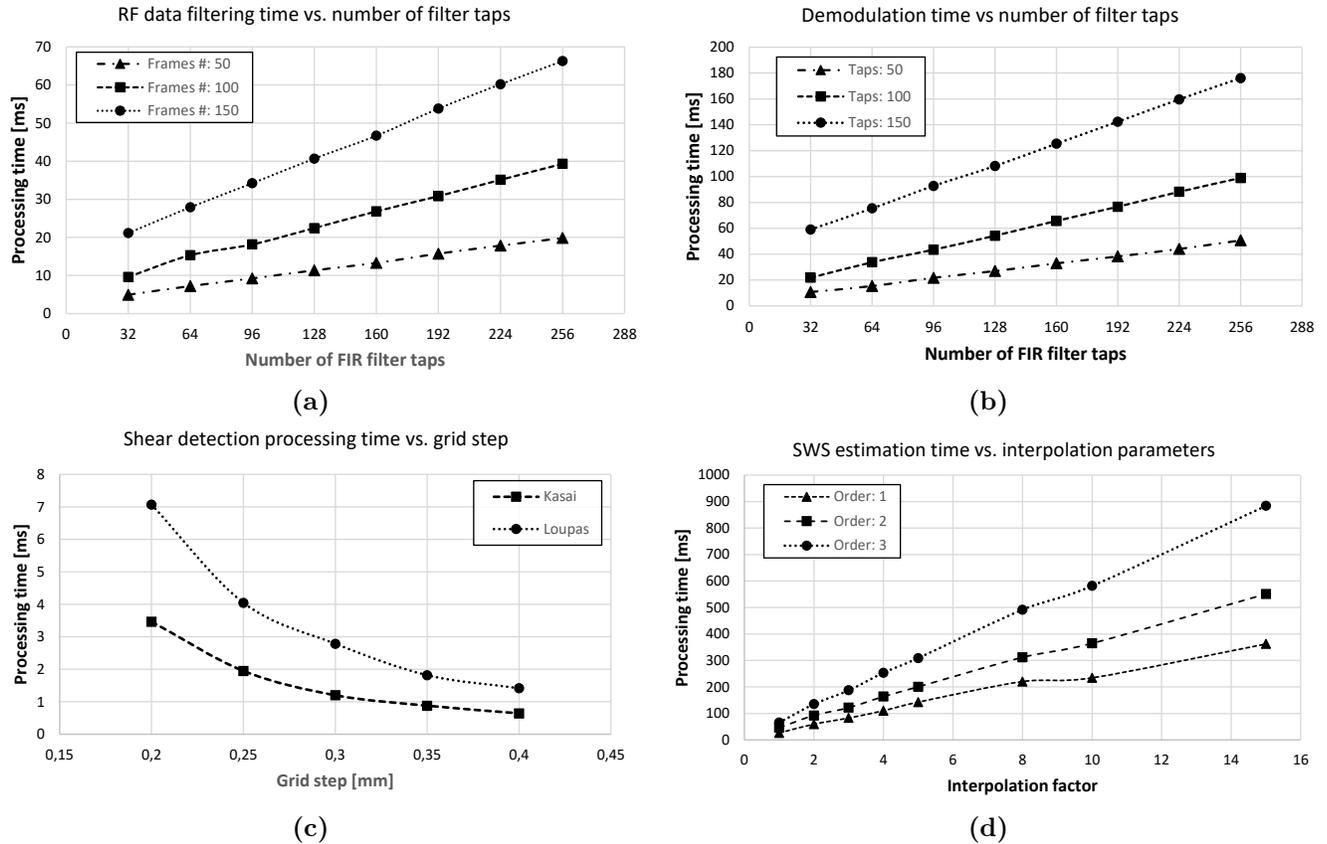
Figure 4.21: Evaluation of the impact of selected parameters on the data processing time by individual algorithm steps. (a) RF data filtering time for various numbers of FIR filter taps and different numbers of frames. (b) Similar characteristics as in (a) but for the demodulation stage. (c) Shear wave detection time versus grid step and the algorithm version. (d) SWS estimation processing time for various interpolation factors and the order of spline interpolation. Grid step of 0.2 mm and 100 frames of input dataset were used.

B)   On the reconstruction parameters tuning for performance and quality

Although the results obtained for a typical set of reconstruction parameters revealed that the SWS estimation requires the most time to process out of all algorithm stages, it was also shown, that the processing time of each stage is affected strongly by both input dataset size and the reconstruction parameters values. Regarding dataset size, at a given shear wave tracking PRF, the number of frames to be processed is defined by the time of shear wave propagation through the FOV, which depends on the medium stiffness. For stiffer mediums, fewer frames may be required to reconstruct the stiffness map in the entire FOV. The obtained results have shown a linear relationship between processing time and the number of frames and this knowledge can be used to predict achievable imaging frame rates for a given medium stiffness. The rest of the evaluated parameters can be configured by the user to achieve the desired performance at the cost of potential image quality loss. It was found, that the key parameters that affect the processing time are the lengths of FIR filters in RF filtering and demodulation stages, the beamforming grid step which defines the computational problem size for all following stages, the shear wave detection method, and interpolation parameters for SWS estimation. For instance, using the following set of parameters: 32 taps of FIR

filters in both RF filtering and demodulation, 0.4 mm beamforming grid step, Kasai's method of shear wave detection, and 2nd order interpolation with a factor of 2 for SWS estimation results in the total processing time of 77.7 ms. In contrast, using 256 taps, 0.2 mm beamforming grid step, Loupas' method of shear wave detection, and 2nd order interpolation with a factor of 15 for SWS estimation results in the total processing time of 751.5 ms. Assuming data acquisition time of 20 ms the first case would achieve a frame rate of around 10 fps, while the second one is only 0.13 fps, which shows a big difference. This example demonstrates the necessity of tuning the reconstruction parameters to find a balance between imaging quality, frame rate, and power consumption for a given application. Finding a reasonable combination of reconstruction parameters requires understanding how each parameter affects the final image quality in terms of resolution, accuracy, and bias. Although an analysis of all these relationships is too extensive to be held in this study, two reconstruction parameters affecting the processing time the most — the beamforming grid step and interpolation parameters for SWS estimation — are briefly evaluated below.

C) Considerations on beamforming grid selection

Fig. 4.22 presents the effect of various grid steps used at the stage of beamforming on the final SWS maps of homogeneous and heterogeneous regions of the phantom. To obtain the images, a similar SSI method was used as described in section 4.2 for the purpose of system validation. One difference was that only two successive pushin beams were used instead of 3 as previously. Beams were focused at depths of 25 and 40 mm, and used 42 and 66 elements of the probe, respectively. Four different grid steps were used: 0.2, 0.4, 0.6, and 0.8 mm. In the case of homogeneous region images (Fig. 4.22(a)), it is hard to spot differences. There are some differences visible at artifacts on shallow depths and also at the bottom of the images and also correct estimation reaches higher depths for finer grids. Nevertheless, quantitative analysis of the estimated values did not reveal any significant differences between images in terms of SNR within an artifact-free ROI. The impact of the beamforming grid step is more apparent in the case of the inclusion SWS maps (Fig. 4.22(b)). Individual pixels of the image become more visible along the grid gets more coarse.

It should be noted, that setting the beamforming grid step of 0.8 mm results in a dataset that contains 16x fewer pixels than when using a grid step of 0.2 mm. Such selection can remarkably reduce the processing time of all subsequent algorithm stages. However, as shown, it may have a negative impact on the resulting images. Typically, a beamforming grid step is picked according to the ultrasound pulse wavelength λ. I.e. for a 4.4 MHz, one cycle pulse and assuming 1540 m/s speed of sound it is common to set the grid step to $\lambda/2 = 0.175$ mm. The B-mode images that are usually captured along with SWS maps would possibly still require the $\lambda/2$ grid step, but the 2-D SWE image reconstruction can be done in an independent process. One should consider the expected resolution of the SWS maps to pick the grid step size for reconstruction. As mentioned earlier in section 2.5.3, the resolution of the 2-D SWE method depends on many factors, while shear wavelength and SWS estimation lateral size have the biggest impact. As mentioned before, shear wavelength can be estimated to be around 2.35 cm in this case, while the SWS estimation lateral kernel size used here was 4 mm. For reference, Fig.4.22(c) shows the effect of using different lateral kernel size (2 mm). One can spot, that individual pixels are more apparent when using a wider kernel. For a kernel of 4 mm, the pixelation is evident for the beamforming grid of 0.8 mm, whereas
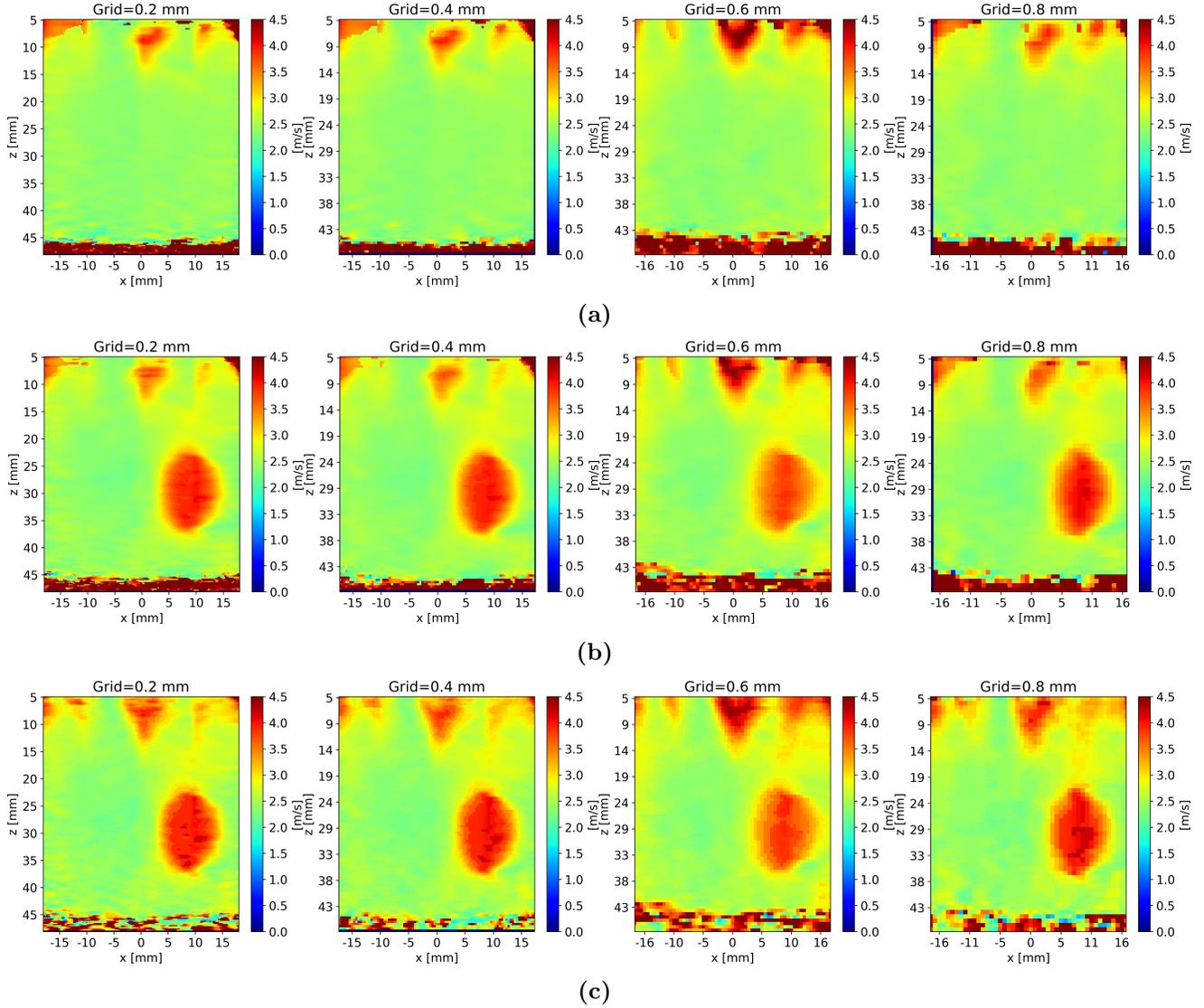
Figure 4.22: SWS maps reconstructed using various grid steps at the stage of beamforming (0.2, 0.4, 0.6, and 0.8 mm, shown left to right, respectively). All maps within each row were reconstructed from the same dataset and used the same (default) reconstruction parameters for all other stages. Images were not post-processed (no smoothing). (a): Homogeneous region (nominal SWS of 2.36 m/s) (b): Inclusion of nominal stiffness of 70.9 kPa and 10.4 mm, reconstructed with the SWS estimation lateral kernel size of 4 mm. (c): Same as in (b), but reconstructed with the SWS estimation lateral kernel size of 2 mm.

for a 2 mm kernel, it is visible for the image obtained with a grid step of 0.6 mm. Thus, in this application, a grid step of lateral kernel size divided by 5 was enough to avoid the pixelation effect.

To summarize, although the 0.2 mm grid step was used by default in this work, it seems reasonable to consider coarsening the grid in applications where processing time or power consumption are key factors, since the image quality degradation may be acceptable or even negligible, especially after smoothing using some kind of image filter.

D)   *Considerations on interpolation parameters for SWS estimation*

Fig. 4.23(a–c) show SWS maps of the homogeneous region of the phantom reconstructed with various interpolation factors (no interpolation, interpolation with a factor of 5, and 15) at the stage of SWS estimation. The SSI method as mentioned in the previous section was used here to obtain the images. Although images look similar, closer inspection reveals big regions of the same estimated value in the case of the image obtained with no interpolation. Differences between Fig. 4.23(b) and Fig. 4.23(c) is not that evident, however, analysis of estimates values distribution within manually selected ROI show that the higher was interpolation factor used, the estimates related to pixels were distributed over a higher number of values. It is expected behavior since — as shown in Eq. 2.5 in Chapter 2 — in the correlation-based ToF SWS estimation method used here the estimates are quantized and the number of possible levels is defined by the tracking sequence FRI. This resolution can be boosted by interpolating the displacement profiles in a slow-time domain. Namely, the SWS estimation resolution $1/\Delta v_{sw}$ is given by:

$$\frac{1}{\Delta v_{sw}} = \frac{FRI}{d\Delta x \cdot I_F},$$

(4.6)

where d lateral estimation kernel size Δx is the pixel size, and I_F is the interpolation factor. In the case of homogeneous regions imaging poor SWS estimation resolution may not be an important issue, since in this case usually a single estimate is obtained by averaging estimates in a big ROI. In this experiment a set of interpolation factors was used: 1, 2, 3, 5, 10, and 15 and an average SWS within the ROI was found to be 2.39 m/s (0.04 m/s bias) for no interpolation case, whereas for other cases the average SWS was found between 2.351–2.359 m/s (bias between 0.001 and 0.1 m/s).

The problem of insufficient SWS estimation resolution may be more severe in the case of imaging inclusions. The SWS maps obtained in a similar experiment to those described above but for heterogeneous region imaging are presented in Fig. 4.24. Images show SWS maps of inclusion of nominal SWS equal to 3.36 m/s and diameter of 16.7 mm. Again, images obtained with small interpolation factors (1, 2, and 3) contain big regions of the same SWS estimate values. Such images can lead to a conclusion, that inclusion is not homogeneous inside and there are stiffer parts and also lead to biased inclusion stiffness assessment. This effect can be considered as an artifact resulting from insufficient interpolation at the stage of SWS estimation. It should be expected, that the smaller and stiffer the inclusion, the effect will manifest stronger.

Regarding spline interpolation order, reconstruction of SWS maps with the same interpolation factor but different interpolation orders (0–3) have shown, that homogeneous regions and inclusions of various stiffnesses obtained for orders in a range of 1–5 were not statistically different when analyzing mean SWS estimates in a ROI or variance of the estimates. Example SWS maps obtained with various interpolation orders are shown in
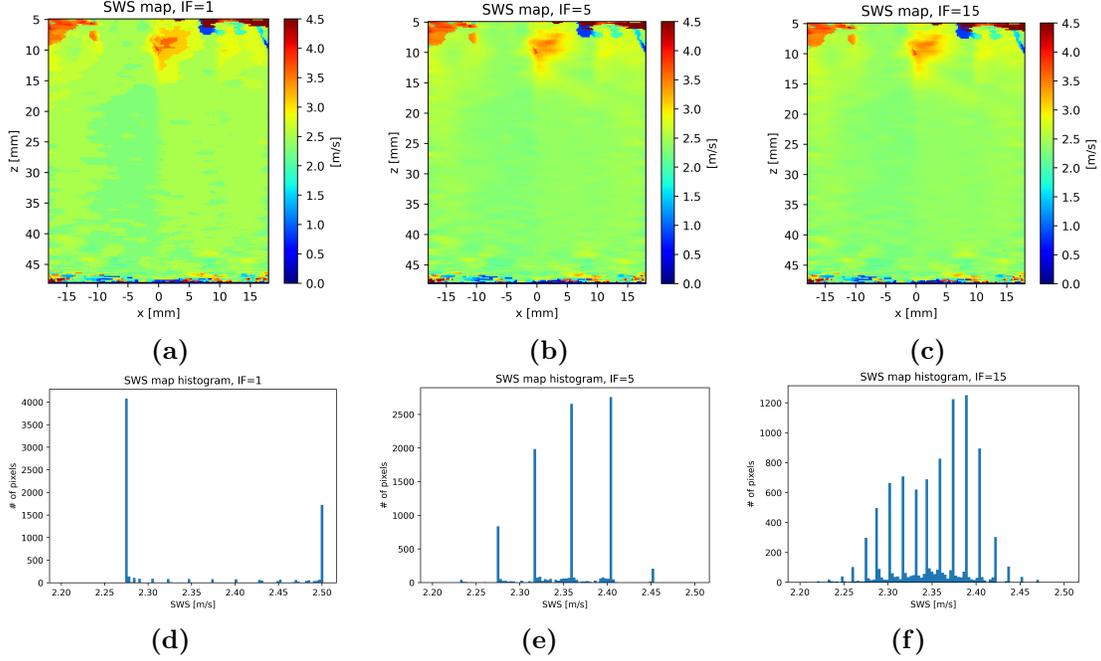
Figure 4.23: Homogeneous phantom (nominal SWS of 2.36 m/s) SWS maps obtained with various interpolation factors used at the stage of SWS estimation. (a): 1 (no interpolation), (b): Interpolated by a factor of 5, and 15 (c). All maps were reconstructed from the same dataset and used the same (default) parameters for all other stages. (d) to (f): histograms of SWS estimates within manually selected, artifact-free 2 x 2 cm ROI in the center part of the corresponding images ((a–c)), respectively. Each histogram consists of 81 bins covering the range 2.2–2.5 m/s.

Fig.4.25. As expected, interpolation of order 0 (equivalent to nearest-neighbor interpolation) exhibited the same effect as images obtained with no interpolation, no matter what interpolation factor was used. Therefore, it should be concluded, that spline interpolation of order 1 (linear interpolation) is generally sufficient in this application.

In summary, the interpolation factor should be picked carefully. On one hand, picking a high factor can lead to an unnecessary processing time increase, from the other hand, picking too low factor can lead to strong artifacts. Some of these artifacts could be up to some extent removed by the post-processing (i.e. median filtering), nevertheless, it is better to avoid original information distortion if it is possible. In general, it seems reasonable to use an interpolation factor that sets SWS estimation resolution being comparable to expected estimation precision. However, one should note, that SWS estimation resolution depends also on tracking sequence FRI and estimation lateral kernel size (see Eq.4.6). The expected precision is also dependent on many parameters, and in particular, it also depends on medium stiffness. Therefore an interpolation factor should be picked individually for each application.

E) *Processing performance comparison with other systems*

Due to the significant impact of acquisition and reconstruction parameters on data processing time, as demonstrated by the presented results, comparing the presented imaging system solution performance to commercial systems is not straightforward. As shown, if considering the computational performance only, the system proposed in this work can potentially provide 2-D elasticity maps with an image refresh rate of up to 10 fps, de-
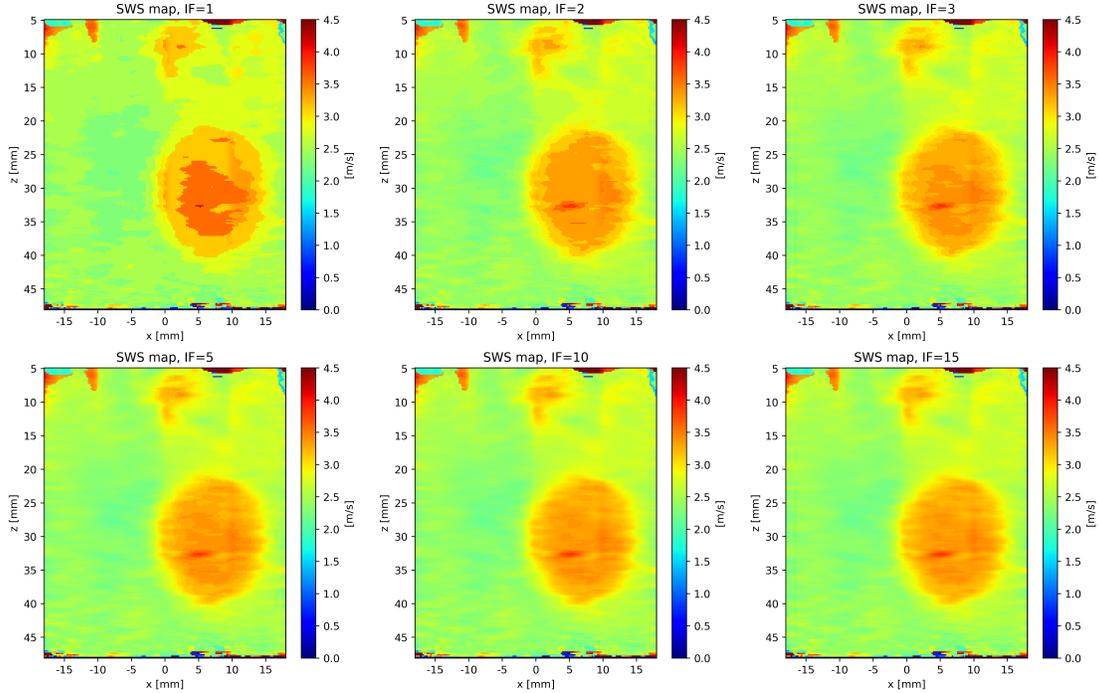
Figure 4.24: Inclusion of nominal SWS equal 3.36 m/s and 10.4 mm diameter SWS maps obtained with various interpolation factors used at the stage of SWS estimation. Top row, left-to-right: 1 (no interpolation), 2, and 3. Bottom row: 5,10, and 15. All maps were reconstructed from the same dataset and using the same (default) parameters for all other stages.

pending on reconstruction parameters settings. However, as pointed out in section 3.4, operating at such a high frame rate can result in excessive imaging system and probe heating and may be a cause of exceeding regulatory acoustic output safety levels. In order to achieve such high frame rates in practical applications one would need to limit the system output power by reducing push duration, limiting the shear wave detection tracking frame rate, or reducing the ROI size. For example, on the General Electric LOGIQ E9/E10/10s scanners, the acquisition time is typically on the order of 100 ms, while the cooling time is typically 1–3 seconds and is the dominant factor in limiting the frame-rate [201]. The pioneer in the field of 2-D SWE-capable imaging systems, Supersonic Imagine (Aix-en-Provence, France), reports for their flagship *Aixplorer Ultimate* scanner the maximum 2-DSWE frame-rate of 3 fps [202]. Among research systems, there are many articles reporting an off-line reconstruction approach for 2-D SWE procedures using a Verasonics scanner, including a paper with guidelines on implementing this modality using this system [123]. However, by the time of writing this text, there are no reports available on the implementation of real-time 2-D SWE imaging on an ultrasound research system.

F)   *Future directions*

In the context of achieving a sufficient image refresh rate, the computational performance of the developed implementation appears to be adequate. However, further optimizations aimed at reducing processing time may be required in cases where the power consumption of the system needs to be limited. Further data processing speed-up
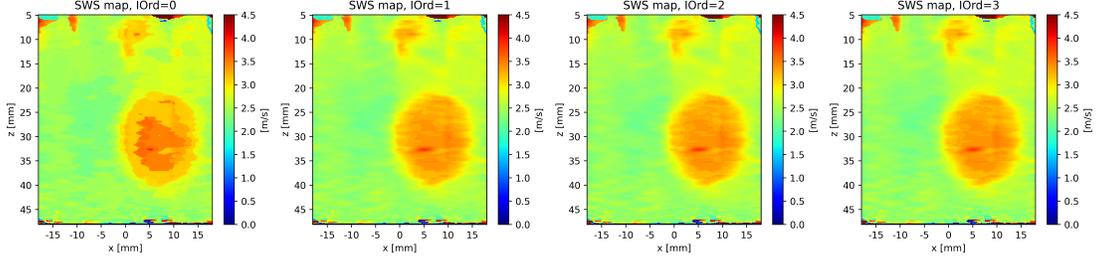
Figure 4.25: Inclusion of nominal SWS equal 3.36 m/s and 10.4 mm diameter SWS maps obtained with various interpolation factors used at the stage of SWS estimation. Left-to-right: 0 (no interpolation), 2, and 3. All maps were reconstructed from the same dataset and using the same (default) parameters for all other stages.

may be achievable by leveraging advanced techniques to optimize algorithm code for GPU execution, like methods to limit the global GPU memory traffic (memory tiling, coalescing, effective usage of shared and constant memory), CUDA streams usage, or other [185]. Nevertheless, this is beyond the scope of this thesis and such works may be planned for the future.

4.3.4 Data processing optimization for reduced power operation

One would tend to reduce data processing time to increase the achievable imaging frame rate. Another reason to reduce processing time can be to decrease the power consumption of the system. This is particularly important in the point-of-care device class, where power consumption plays a vital role. To determine the contribution of data processing to the imaging system's power budget required to obtain a single image, a simple energy model is proposed below.

A)   Energy model

The total electrical energy E_{TOTAL} required to obtain a single image in the 2-D SWE method can be modeled as consisting of four components:

$$E_{TOTAL} = E_{PUSH} + E_{PWI} + E_{SYS} + E_{REC}, \tag{4.7}$$

where:

- E_{PUSH} — is the energy required to generate the pushing sequence;

- E_{PWI} — is the energy required to generate transmit pulses for shear wave tracking;

- E_{SYS} — is the energy consumed by the imaging system during data acquisition, including data transfer;

- E_{REC} — is the energy consumed by the data computing engine during data processing for image reconstruction.

An average power of the system can be found by dividing E_{TOTAL} by the frame repetition interval T_{FRI}: $P_{TOTAL} = E_{TOTAL}/T_{FRI}$. Power components P_{PUSH}, P_{PWI}, P_{SYS}, and

121

P_{REC} can be found in the same way. The subsequent components of the model can be found as follows:

$$E_{PUSH} = 4C_t \cdot U^2 \cdot \sum_{i=1}^{N_B} a_i \cdot T_i \cdot f_i, \tag{4.8}$$

where C_t is the transducer element electrical capacitance, U is the excitation voltage (single rail voltage level), N_B is the number of beams of a push pulse, a_i is the number of transducer elements for ith beam, T_i is the ith push duration, and f_i is the ith push pulse center frequency.

$$E_{PWI} = 4C_t \cdot U^2 \cdot n_{ch} \cdot n_{cycles} \cdot n_{frames}, \tag{4.9}$$

where n_ch is the number of active transmit channels per insonification, n_{cycles} is number of cycles of the transmit pulse, and n_{frames} is the number of acquisition frames.

$$E_{SYS} = P_{sys_A} \cdot (T_{acq} + T_t) + P_{sys_B} \cdot (T_{FRI} - T_{acq} - T_t), \tag{4.10}$$

where P_{sys_A} and P_{sys_B} are the average acquisition system power consumptions during active operation (data acquisition and transfer) and during idle state, respectively, T_{acq} is total acquisition time and T_t is data transfer time, and T_{FRI} is the frame repetition interval.

$$E_{REC} = \sum_{i=1}^{N} P_i \cdot T_i + P_{standby} \cdot \left(T_{FRI} - \sum_{i=1}^{N} T_i \right), \tag{4.11}$$

where P_i is the average power of the computational unit during data processing of ith algorithm step, T_i processing time for this step, and $P_{standby}$ is the average power of the computational unit in idle state.

B) Evaluated scenario

Let's consider an example imaging scenario as proposed in real-time implementation earlier in this thesis (see section 3.4.6).

- Push generation sequence:

    - two parallel pushing beams generated in a comb as in the CUSE method,
    - at both lateral locations two subsequent beams are generated focused at two consecutive depths (as in the SSI method), using 42 and 63 transducer elements, respectively,
    - push frequency 4.4 MHz (applies to each individual beam),
    - push duration 400 μs (applies to each individual beam).

- Tracking sequence:

    - Full aperture of 128 transducer elements,
    - Transmit pulse: 4.4 MHz, 3 cycles.
    - Tracking PRF of 5 kHz,
    - Collecting 100 frames, each consisting of 4200 samples from each channel at a sampling frequency of 65 MHz.

The values for the remaining variables were obtained as per the following: high-voltage supply rails of ± 70 V, C_t of 200 pF (measured for ATL L7-4 probe element using impedance bridge), P_{sys_A} and P_{sys_B} equal 50 W (nominal for two us4OEM modules) and 5 W (module power-down assumed), respectively, imaging frame rate of 2 fps, and data transfer throughput of 2.4 GB/s resulting in T_t of 42 ms. Average GPU power during active processing was found to be $\sim$74 W with very low differences when performing consecutive reconstruction algorithm stages. Average GPU power consumption in standby $P_{standby}$ was measured to be $\sim$20 W. The GPU power was measured using the Nvidia Nsight Systems performance analysis tool (Nvidia, Santa Clara, CA, USA). Unless otherwise mentioned, the total processing time was assumed to be 348 ms as found in the previous section for the default set of reconstruction parameters.

C) Results

The power distribution between imaging process components is shown in Fig. 4.26(a). The biggest contribution (57.6 W) was brought by data processing, which is 82 % of total power, while the second biggest was acquisition system operation, responsible for 10.2 W (14.3%). In the analyzed scenario power required for push sequence generation was equal to 2.9 W (4.1%), while the power consumed by shear wave tracking was found negligible (0.3 W, 0.04%). This shows, that power consumption optimization of the system should be started from the data processing component. An obvious strategy would rely on reducing data processing time. Fig. 4.26(b) demonstrates how the total power consumption changes along with data processing time. For instance, by reducing processing time from 350 ms down to 100 ms, the total power would decrease from 71.2 W to 44.2 W (-38%). Still, however, image reconstruction would be responsible for almost 70 % of total power.
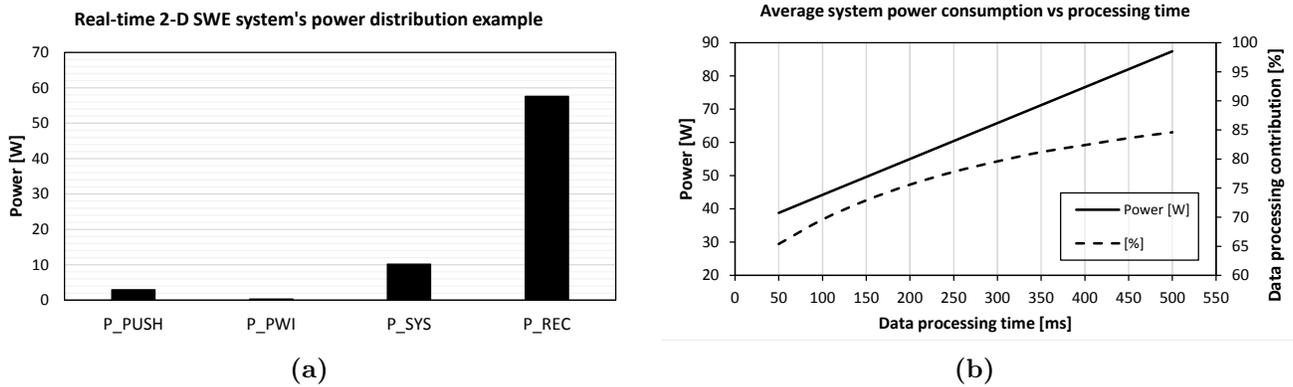
Figure 4.26: Power usage in analyzed 2-D SWE imaging scenario. TX/RX sequences and all parameters are described in the text. (a): System's power distribution between imaging process components: push generation sequence P_{PUSH}, tracking sequence P_{PWI}, data acquisition system operation P_{SYS}, data processing for image reconstruction P_{REC}. (b): Solid line: average system total power consumption for various data processing times. Refers to the axis on the left. Dashed line: the percentage that represents the power required for processing data for image reconstruction in relation to the total power. Refers to the axis on the right.

D) Discussion and conclusions

The presented analysis has some limitations. Firstly, the usage of Nvidia RTX 3080 GPU was assumed to be used as a processing engine. It is a powerful GPU, not necessarily

optimized in terms of power. Anyway, this particular device was used in the model, since it was a part of the setup in this thesis. In a truly power-optimized system, one would exploit a portable GPU module, more suitable for portable devices with strict power constraints. However, it is hard to predict, what would be the result of migrating the data processing to such GPU with respect to processing time. Migration of developed data processing to the portable computing back-end such as Nvidia Jetson Xavier is one of the future directions following this thesis.

In summary, it was shown, that the data processing for the image reconstruction process is the most power-intensive part of the developed system. Reducing the processing workload to minimize the processing time and, consequently, the consumed power seems to be the most obvious approach for power optimization of the system. One may adjust the reconstruction parameters affecting the processing time the most to find a balance between performance, power/energy consumption, and imaging quality. For instance, by combining results from this and the previous section, one can find, that in the analyzed case by using a reduced number of FIR filter taps to 64 (instead of 236 and 128, for RF data filtering and demodulation stages, respectively), beamforming grid step of 0.4 mm (instead of the default 0.2 mm), interpolation factor of 3 (instead of default 5), and interpolation order of 1 (instead of default 2), one can expect to reduce data processing time down to around 100 ms. As found above, in the analyzed system, that would lead to a power consumption decrease of 38%. However, as shown in the previous section, the effect of such parameters change on the imaging quality should be evaluated in a given application.

Regarding other components of the 2-D SWE imaging process, the conducted analysis has shown, that the acquisition system power consumption would be a dominating factor only if one would be able to decrease computing back-end power usage by a factor of 6, while the acquisition parameters of push generation or tracking sequences seem to have negligible effect. This leads to the conclusion, that the power optimization of the 2-D SWE imaging implementation should be focused on improvements in the data processing and in hardware design.

Chapter 5

Summary

This chapter concludes the thesis by summarizing the presented works and discussing the achieved results. Summaries of the contributions in four main areas are given: theoretical works, hardware development, software development, and the experimental part. Next, the thesis limitations are listed and discussed, and future directions are proposed as well. Conclusions are given at the end.

5.1   Thesis summary

The main goal of this thesis has been achieved — the 2-D shear wave elastography method has been successfully implemented. In the presented approach the us4R-lite ultrasound research system was used as a base hardware platform that the 2-D SWE was developed over. Multiple enhancements were required in order to provide system capabilities allowing data acquisition and data processing for this complex technique. The efforts focused on both hardware and software parts of the system resulted in the successful development of data acquisition and data processing procedures of the performance providing real-time 2-D tissue stiffness maps capture.

Successful implementation of such a sophisticated technique as 2-D SWE using a portable, low-cost ultrasound research system is the main achievement of this thesis. To the author's best knowledge, it is the first real-time system solution implementing 2-D SWE mode entirely developed in Poland, and the smallest real-time 2-D SWE capable system in the World. This achievement can be considered a first step toward transferring this technology to truly portable or even handheld systems.

The experimental part of the thesis presented the original research aimed to evaluate and optimize various aspects of the 2-D SWE technique implementation to make it more suitable for point-of-care class devices. The conducted research focused on three areas:

- evaluation of 2-D SWE procedure safety in terms of acoustic output related to push pulse generation;

- developed system validation — including data capture at various stages of processing, assessment of imaging quality, and exposing the typical problems in the 2-D SWE method limiting imaging quality;

- evaluation of 2-D SWE image reconstruction algorithm performance, including reconstruction parameters adjustment for a balance between processing time, power consumption, and imaging quality;

The thesis formulated at the beginning of the manuscript was proven true; the 2-D ultrasound shear wave elastography technique was practically implemented using a point-of-care device class by exploiting a GPU-accelerated, software-defined ultrasound approach. In the proposed system solution the entire data processing was carried out within the software to maximize flexibility, while the processing performance for real-time imaging was achieved by utilizing parallel processing with GPU. It was also shown, that the developed system — even at that small size and with limited hardware — exhibited comparable imaging quality with respect to bias and frame rate, and only slightly worse performance in terms of spatial resolution and accuracy, in comparison to cart-based commercial systems. Nevertheless, it should be noted, that the comparison was based on literature research only due to the lack of access to these systems to conduct a direct comparison. Moreover, the 2-D SWE imaging quality depends on many parameters related to acquisition schemes or data processing which also affect frame rate and power consumption. Therefore, a fair comparison should include an analysis of the factor of power, which was not possible to derive from literature research.

The overall methodology of works targeted to achieve the aims of the thesis was described in five chapters of this manuscript. In the first, introductory chapter 1, to build the theoretical background, the physical principles of the SWE were discussed and known ultrasonic tissue elasticity imaging methods were explored. The key strengths and disadvantages of each method were identified which allowed to evaluate these methods in terms of their suitability for the implementation using a portable imaging system. In the theoretical chapter 2, the data acquisition and data processing methods for the 2-D SWE technique were reviewed in detail. This analysis provided a theoretical background for the transmit and receive procedures and a selection of methods for image reconstruction algorithm to be implemented. It also allowed the definition of technical requirements to be met for the practical implementation of this method. The processes of hardware and software developments for the implementation of the 2-D SWE using the us4R-lite research system were described in chapter 3, including the considerations on methods, optimization, and tool selection for providing real-time imaging capabilities. In chapter 4 the developed system was validated in a series of experiments held using an elastography phantom. The achieved performance was compared to commercial systems. In order to reduce the power consumption of the processing back-end, the reconstruction algorithm parameters were evaluated in terms of their effect on processing time and imaging quality. Imaging procedure safety was evaluated by hydrophone-based acoustic output measurements.

5.2 Thesis contributions

As a supplement to the concise summary above, a more detailed description of the contributions in three main areas of the thesis is given below. Each section also discusses the significance of achievements in each area. Contributions that were used and were not brought by the thesis author are also listed at the end of this section.

5.2.1 Theoretical contributions

The main theoretical contribution of this thesis is the detailed literature review on the 2-D SWE method provided in chapter 2. Several aspects of this technique were addressed, including the generation of push pulses and pushing sequences, data acquisition

schemes, approaches in data processing for image reconstruction, safety considerations, and elaboration on imaging quality in terms of bias, uncertainty, and resolution. Except for well-established techniques, novel methods in all areas mentioned above were briefly discussed. Although overviews of ultrasound-based tissue elasticity imaging methods were written by multiple authors ([18], [28], [42], [203]), to the best knowledge of the author, the contents of chapter 2 is the first such study encompassing discussion of all relevant aspects of 2-D SWE in concise form and one place.

Due to the practical nature of this thesis, the overall methodology relied on using an efficient combination of known methods rather than proposing novel theoretical concepts. Nevertheless, a review of techniques in the 2-D SWE method presented in the theoretical part of this dissertation can be found useful for researchers or engineers interested in the implementation of this method or similar.

5.2.2 Hardware developments

The main hardware contribution of this thesis is the development of the Transmit Push Beamformer Board — TXPB-256 (Fig. 3.3) — a dedicated 256-channel transmitter subsystem for the us4R-lite research system. It was designed in order to increase the transmit capabilities of the us4R-lite platform to meet requirements of the 2-D SWE method. Designed in the form of an extension module that could be easily integrated with the us4R-lite system, it utilized the new generation transmit pulsers ICs as the transmit excitation devices. This development was the key element that allowed the practical implementation of the SWE technique since it enabled the generation of high-energy excitation for push pulse generation that induces shear waves in the medium. By providing push pulse generation of any pattern, the us4R-lite system enhanced with TXPB-256 can be used to develop and evaluate other techniques based on ARFI excitation.

Another important contribution of this thesis is the firmware design of the TXPB-256 onboard FPGA device. In this module, the FPGA acted as a high-speed control unit, which reconfigured the TX pulsers dynamically during runtime, depending on a pre-defined transmit plan configured by software for desired imaging scenario. Within configurable FPGA architecture, several functional IP blocks have been designed, each intended to perform different tasks. Each of the IP blocks has been written in VHDL to be easily transferable between FPGA devices and across the projects.

Although the TXPB-256 module was designed specifically for 2-D SWE method development, its high configurability of transmit parameters provides flexibility in designing transmit patterns that can be found useful in many other applications. Indeed, by the time of writing this text, the transmit subsystem design used on TXPB-256 is under migration to the new version of the us4R-lite platform. Some firmware IP blocks developed for the purpose of this thesis have been already exploited in currently available us4R-lite systems. Moreover, the main part of the developed FPGA firmware design was successfully migrated to the us4R-lite FPGAs on the new generation of us4R-lite systems (us4R-lite+). Thus, these new generations of this research system will support dynamic elastography applications by default.

5.2.3 Software developments

The main software contribution of this thesis is the implementation of the complete image reconstruction algorithm for the 2-D SWE method for stiffness map reconstruction using

acquired RF channel data as an input. A dedicated multi-step algorithm was developed in two versions.

The first one was the offline version implemented MATLAB environment. This implementation allowed off-line image reconstruction for algorithm validation, debug purposes, and rapid development of new data processing approaches. The algorithm consisted of the following steps: RF data filtering, demodulation, beamforming, shear wave detection, shear wave motion data filtering, shear wave speed estimation, image compounding, and post-processing. One of the target features of the reconstruction algorithm implementation was to provide its high flexibility. Therefore, the algorithm features many configurable parameters to allow evaluation of their effect on the final image quality, profile power vs. performance trade-off, or to adjust them in a particular application.

In the second implementation, the same algorithm was adapted for real-time execution in GPU-accelerated implementation using Python and CUDA C. Computational-intensive algorithm steps were mapped into computing kernels to be executed on GPU. Each kernel involved careful problem decomposition for parallel processing to effectively exploit underlying execution by GPU hardware. These efforts resulted in the real-time performance of the algorithm, reaching frame rates in the range of up to around 9 fps, depending on acquisition type and reconstruction parameters settings.

Real-time implementation was integrated with the hardware by a set of tools allowing straightforward acquisition configuration for some common 2-D SWE imaging scenarios. In addition, due to the modular architecture of both implementations, they can be easily used as a base for the development of novel data processing procedures in any of the image reconstruction steps. Instead of writing the complete algorithm from scratch, a single kernel can be modified or replaced by a new one. Moreover, with the scalable nature of CUDA architecture, the GPU-accelerated, real-time algorithm design is easily transferable between GPU devices, for example, from high-performance to low-power, mobile GPU. For this reason, the developed hardware-software framework can be used as a platform for an evaluation of elasticity imaging methods or feasibility studies. Examples of such an approach in the work published by the author [80].

5.2.4 Experimental part of the thesis

In the first experiment, a dedicated setup was built to characterize the acoustic output of the developed system during push pulse generation. Hydrophone-based measurements of the pressure generated by two pushing beams of various focal configurations have shown, that the generated intensities were within regulatory limits of MI and I_{SPTA}. However, it was shown, that pushing patterns for specific applications should be designed carefully, especially in real-time operation with high push PRF, since increasing the number of push beams or excitation strength in terms of push duration or magnitude can lead to exceeding the safety thresholds.

For validation purposes, the developed system operation was verified experimentally using an industry-standard elastography phantom. Homogeneous and heterogeneous phantom experiments were held to evaluate imaging bias, accuracy, and resolution. In homogeneous region (nominal 2.36 m/s SWS of the material) imaging using the SSI procedure SWS estimation bias was found to be $< 5\%$ and the estimates' variance in a ROI between 0.06 to 0.13 m/s, what corresponds to SNR in a range of 29–32 dB. Single pixels exhibited an average estimation variance of 0.044 m/s when analyzing the same pixels

from multiple repeated acquisitions. Regarding imaging of inclusions, CNR (9.2–17 dB), bias (-23.2 to 32.4%), SWS estimation variance within inclusions (0.06–0.28 m/s), and other quality metrics depended strongly on inclusion stiffness and diameter. The system was able to detect inclusions of diameters $\geq$6.5 mm. In all cases, the imaging depth was limited to around 40 mm. In the original results obtained at various stages of data processing some typical problems were exposed that are limiting the 2-D SWE imaging quality, namely: non-uniform shear wave energy distribution, shear wave attenuation, low shear wave bandwidth, limited energy transfer at greater depths, speckle bias, bias patterns, limited resolution and significant SWS estimation bias of inclusions of small diameter. Identification of these problems and their significance can lead to novel methods aimed at solving them to improve imaging performance. This topic was up to some extent addressed later in this thesis.

The overall results obtained in the third part of the experimental portion of the dissertation demonstrated that the developed system solution exhibited a level of performance that enabled real-time 2-D SWE imaging through the utilization of parallel processing using a GPU. With a default set of reconstruction parameters, the full SWS map reconstruction was performed in 348 ms (2.85 frames/s). However, it was also shown, that the processing time of each algorithm stage was affected strongly by both input dataset size and the reconstruction parameters values. These parameters could be configured to achieve the desired performance at the cost of potential image quality loss. All relevant reconstruction parameters were evaluated in terms of their effect on processing time. The parameters of strongest impact were identified to beamforming grid step and interpolation parameters for SWS estimation. The effect of both parameters on resulting image characteristics was discussed, which allowed to formulate conclusions on adjusting these parameters' values in a given application. Then, in order to evaluate factors affecting the power consumption of the 2-D SWE image capture process the energy model was proposed. Modeling revealed, that in the developed system, the data processing for image reconstruction brought the biggest contribution to the system's power budget. It was also shown, that adjustment of reconstruction parameters values can lead to significant power savings.

5.3 Thesis limitations

Due to finite scope or lack of resources, the presented dissertation has some limitations. The following limitations can be listed as the most obvious:

- The imaging capabilities of the developed system were not directly compared experimentally with any commercial system equipped with 2-D SWE mode. Imaging of the same phantom using two systems could lead to a fair comparison and a solid reference for the designed system performance. Unfortunately, the author had no access to such a commercial system. Therefore, the comparison was made only based on results published by other authors using commercial systems.

- In the evaluation of the system 2-D SWE imaging bias and stiffness estimation precision only one homogeneous phantom was used, with a single nominal stiffness value (16.7 kPa). The same material was a background for the inclusions in heterogeneous experiments. Bringing another variable for evaluation using various stiffnesses in both cases would be beneficial as it would possibly expose more char-

acteristics of the imaging system. Nevertheless, due to the high cost of such phantoms, only one was available for this study. Since guidelines for designing custom tissue-mimicking phantoms have been published by multiple authors (e.g. [204]), the usage of self-prepared phantoms was possible (as, for example, in [59], [114], [196]. However, phantoms prepared this way would require characterization before use by mechanical methods or MRI. Due to the lack of experience in both phantom fabrication and characterization, it was decided to limit experiments to only one phantom that was calibrated and available.

- The experimental validation did not include any *in-vivo* study. However, this topic was decided to be out of scope due to no possibility of comparing obtained results with reference commercial systems. Nevertheless, such works can be one of the future research directions.

5.4 Future directions

The open architecture ultrasound imaging system enhanced in this thesis to implement the 2-D SWE method has many features that make it an attractive hardware platform to carry out various types of studies in the field of dynamic elastography. It is highly flexible in terms of TX/RX sequence configuration and allows any manipulation of captured data, including parallel processing using the algorithm developed in this thesis as a starting point. With the results obtained in this thesis and using the developed framework, the following research topics were found interesting and are considered to be explored in the future:

- Optimizing pushing beam sequence generation to maximize image quality at a given frame rate and power constraints. Adjusting parameters like f-number, push length, push frequency of individual beams, or beam spacing within typically used patterns (SSI, CUSE, F-CUSE) may result in image quality improvement in certain applications. By using the developed system, both measurements and data analysis can be fully scripted and automated.

- Similar research as described above can be held to take data acquisition parameters into consideration, such as: frame rate, shear wave tracking method, and tracking pulses frequency. One possible direction would be to evaluate various data interpolation techniques to maximize frame rate without sacrificing image quality since some groups reported promising results [117], [200], [205]. This approach also could be exploited in receive channel-limited systems. Another interesting topic would involve using coded excitation [121] or harmonic imaging [120], [133] to improve shear wave detection SNR at greater depths.

- Incorporate measurement of shear wave speed dispersion to assess information about tissue viscoelasticity. This aim can be realized twofold — by adapting a data acquisition procedure to induce shear waves of varying frequencies, or by utilizing smart data processing strategies, for instance, Local Phase Velocity Imaging (LPVI) method [81], [206].

- Perform ultrasound modes fusion with 2-D SWE method to enhance diagnostic information. One possible application that would potentially benefit from B-mode

imaging, PW Doppler, and 2-D SWE fusion is an arterial wall and blood flow imaging for vascular disease assessment [33], [65].

Another direction that future works can be focused on involves further development of the existing 2-D SWE system solution. Future development may be focused on more tight integration of the new transmitter circuit with the receiver circuit. In the developed solution, the TXPB-256 module comes as an extension board for the us4R-lite system. Circuit from this module can be integrated with the rest of the circuits of the us4OEM module to minimize device size. As mentioned earlier, at the time of writing this text this is an ongoing process. Moreover, the portability of the system may be increased by embedding the computing back-end within the us4R-lite scanner. At the same time, by utilizing power-optimized GPU for mobile applications the system would become more efficient in terms of power.

5.5 Conclusions

Based on the overall results presented in this thesis, the following conclusions can be formulated:

1. The 2-D shear wave elastography is a complex ultrasound imaging technique featuring strict hardware requirements and requiring high computational performance due to the huge amount of data to be processed and the computationally intensive nature of the image reconstruction algorithm.

2. By careful hardware and software design of the system, the real-time 2-D ultrasound shear wave elastography imaging can be practically implemented using a point-of-care device class by exploiting a GPU-accelerated, software-defined ultrasound approach.

3. Providing real-time imaging capability for the 2-D SWE method is possible through the utilization of parallel processing using a GPU. Based on the results obtained in this work it is justified to expect, that exploiting GPU to the data processing for 2-D SWE methods can reduce the processing time even up to 100x, depending on the degree of data parallelism in the algorithm, optimization techniques usage and number of resources of the GPU.

4. Development of the 2-D SWE method using a portable research platform of an open architecture opens a wide variety of possible research directions that may result in the development of novel methods in areas of data acquisition and data processing. Possible directions include: pushing sequence generation optimization, development of new acquisition sequences, validation of novel approaches in data processing algorithms, or fusion with other ultrasound imaging modalities to enhance diagnostic information in complex applications.

■

Bibliography

[1] Hippocrates, *The Book of Prognostics*, trans. by F. Adams. 399. [Online]. Available: `https://en.wikisource.org/wiki/The_Book_of_Prognostics`.

[2] K. K. Shung, *Diagnostic Ultrasound: imaging and blood flow measurements*. CRC Press, Taylor & Francis Group, 2006, ISBN: 0-8247-4096-3.

[3] J. Etienne, L. Filipczyński, J. Groniowski, *et al.*, *Diagnostyka Ultradźwiękowa w Położnictwie i Chorobach Kobiecych*, L. Filipczyński and I. Roszkowski, Eds. Państwowy Zakład Wydawnictw Lekaraskich, 1977.

[4] T. L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out*. Elsevier, 2014, ISBN: 978-0-12-396487-8.

[5] Siemens. "Imaging: Medical technology." (2022), [Online]. Available: `https://new.siemens.com/global/en/company/about/history/technology/medical-technology/imaging.html` (visited on 05/07/2022).

[6] C. Ligtvoet, L. Rusterborgh, Kappen, and N. Bom, "Real time ultrasonic imaging with a hand-held scanner part I—technical description," *Ultrasound in Medicine & Biology*, vol. 4, no. 2, pp. 91–92, 1978.

[7] H. So, J. Chen, B. Yiu, and A. Yu, "Medical ultrasound imaging: To GPU or not to GPU?" *IEEE Micro*, vol. 31, no. 5, pp. 54–65, 2011.

[8] J. Bercoff, "Ultrafast imaging in biomedical ultrasound," in *Ultrasound Imaging - Medical Applications*, P. O. Minin, Ed. InTech, 2011, ISBN: 978-953-307-279-1. [Online]. Available: `http://www.intechopen.com/books/ultrasoundimaging-medical-applications/ultrafast-ultrasound-imaging`.

[9] E. Boni, A. C. H. Yu, S. Freear, J. A. Jensen, and P. Tortoli, "Ultrasound open platforms for next-generation imaging technique development," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 7, pp. 1078–1092, 2018.

[10] M. Tanter and M. Fink, "Ultrafast imaging in biomedical ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 61, no. 1, pp. 102–119, 2014.

[11] K. Thiele, "Commercializing CMUTs and deep learning to realize worldwide ultrasound democratization," presented at the IEEE International Ultrasonics Symposium (Oct. 10, 2019), Glasgow, Scotland, UK, 2019.

[12] I. O. Wygant, X. Zhuang, D. T. Yeh, *et al.*, "Integration of 2D CMUT arrays with front-end electronics for volumetric ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 55, no. 2, pp. 327–342, 2008.

[13] A. Bhuyan, J. W. Choe, B. C. Lee, *et al.*, "Integrated circuits for volumetric ultrasound imaging with 2-D CMUT arrays," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 7, no. 6, pp. 796–804, 2013.

[14] C. F. Dietrich, A. Goudie, L. Chiorean, *et al.*, "Point of care ultrasound: A WFSUMB position paper," *Ultrasound in Medicine & Biology*, vol. 43, no. 1, pp. 49–58, 2017.

[15] C. A. Andersen, S. Holden, J. Vela, M. S. Rathleff, and M. B. Jensen, "Point-of-care ultrasound in general practice: A systematic review," *Annals of family medicine*, vol. 17, no. 1, pp. 61–69, 2019.

[16] B. Sorensen and S. Hunskaar, "Point-of-care ultrasound in primary care: A systematic review of generalist performed point-of-care ultrasound in unselected populations," *The ultrasound journal*, vol. 11, no. 1, 2019.

[17] L. Lee and J. M. DeCara, "Point-of-care ultrasound," *Current Cardiology Reports*, vol. 17, no. 1, 2020.

[18] A. Sarvazyan, T. J. Hall, M. W. Urban, M. Fatemi, S. R. Aglyamov, and B. S. Garra, "An overview of elastography–an emerging branch of medical imaging," *Current Medical Imaging Reviews*, vol. 7, no. 4, pp. 255–282, 2011.

[19] J. F. Greenleaf, M. Fatemi, and M. Insana, "Selected methods for imaging elastic properties of biological tissues," *Annual Review of Biomedical Engineering*, vol. 5, pp. 57–78, 2003.

[20] K. J. Parker, M. M. Doyley, and D. J. Rubens, "Imaging the elastic properties of tissue: The 20 year perspective," *Physics in Medicine and Biology*, vol. 56, no. 1, pp. 1–29, 2011.

[21] S. J. Kwon and M. K. Jeong, "Advances in ultrasound elasticity imaging," *Biomedical Engineering Letters*, vol. 7, no. 2, pp. 71–79, 2017.

[22] J.-L. Gennisson, "Rheological model-based methods for estimating tissue viscoelasticity," in *Ultrasound Elastography for Biomedical Applications and Medicine*, I. Nenadic, N. Urban, J. Greenleaf, J.-L. Gennisson, M. Bernal, and M. Tanter, Eds. Wiley & Sons, 2019, ISBN: 9781119021551.

[23] T. Shiina, K. R. Nightingale, M. L. Palmeri, *et al.*, "WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 1: Basic principles and terminology," *Ultrasound in Medicine & Biology*, vol. 41, no. 5, pp. 1126–1147, 2015.

[24] J. R. Doherty, G. E. Trahey, K. R. Nightingale, and M. L. Palmeri, "Acoustic radiation force elasticity imaging in diagnostic ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 60, no. 4, pp. 685–701, 2013.

[25] L. V. amd Jean-Luc Gennisson and I. Nenadic, "Continuum mechanics tensor calculus and solutions to wave equations," in *Ultrasound Elastography for Biomedical Applications and Medicine*, I. Nenadic, N. Urban, J. Greenleaf, J.-L. Gennisson, M. Bernal, and M. Tanter, Eds. Wiley & Sons, 2019, ISBN: 9781119021551.

[26] T. Deffieux, G. Montaldo, M. Tanter, and M. Fink, "Shear wave spectroscopy for in vivo quantification of human soft tissues visco-elasticity," *IEEE Transactions on Medical Imaging*, vol. 28, no. 3, pp. 313–322, 2009.

[27] S. Chen, M. W. Urban, C. Pislaru, *et al.*, "Shearwave dispersion ultrasound vibrometry (SDUV) for measuring tissue elasticity and viscosity," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 1, pp. 55–62, 2009.

[28] J. Bamber, D. Cosgrove, C. F. Dietrich, *et al.*, "EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. part 1: Basic principles and technology," *Ultrashall in der Medizin*, vol. 34, no. 2, pp. 169–184, 2013.

[29] J.-L. Gennisson, T. Deffieux, E. Macé, G. Montaldo, and M. F. amd Mickaël Tanter, "Viscoelastic and anisotropic mechanical properties of in vivo muscle tissue assessed by supersonic shear imaging," *Ultrasound in Medicine & Biology*, vol. 36, no. 5, pp. 789–801, 2019.

[30] L. Coutts, J. Bamber, and N. Miller, "Multi-directional in vivo tensile skin stiffness measurement for the design of a reproducible tensile strain elastography protocol," *Skin Research and Technology*, vol. 19, no. 1, pp. 37–44, 2013.

[31] J.-L. Gennisson, N. Grenier, C. Combe, and M. Tanter, "Supersonic shear wave elastography of in vivo pig kidney: Influence of blood pressure, urinary pressure and tissue anisotropy," *Ultrasound in Medicine & Biology*, vol. 38, no. 9, pp. 1559–1567, 2012.

[32] S. S. Leong, J. H. D. Wong, M. N. M. Shah, *et al.*, "Stiffness and anisotropy effect on shear wave elastography: A phantom and in vivo renal study," *Ultrasound in Medicine & Biology*, vol. 36, no. 1, pp. 34–45, 2020.

[33] M. Couade, M. Pernot, C. Prada, *et al.*, "Quantitative assessment of arterial wall biomechanical properties using shear wave imaging," *Ultrasound in Medicine & Biology*, vol. 36, no. 10, pp. 1662–1676, 2010.

[34] G. R. Torr, "The acoustic radiation force," *American Journal of Physics*, vol. 52, no. 5, pp. 402–408, 1984.

[35] K. R. Nightingale, M. L. Palmeri, R. W. Nightingale, and G. E. Trahey, "On the feasibility of remote palpation using acoustic radiation forces," *The Journal of Acoustical Society of America*, vol. 110, no. 1, pp. 625–634, 2001.

[36] K. Nightingale, M. S. Soo, R. Nightingale, and G. Trahey, "Acoustic radiation force impulse imaging: In vivo demonstration of clinical feasibility," *Ultrasound in Medicine & Biology*, vol. 28, no. 2, pp. 227–235, 2002.

[37] T. Sugimoto, S. Ueha, and K. Itoh, "Tissue hardness measurement using the radiation force of focused ultrasound," *Proceedings of IEEE Symposium on Ultrasonics*, 1990.

[38] K. Nightingale, R. Bentley, and G. Trahey, "Observations of tissue response to acoustic radiation force: Opportunities for imaging," *Ultrasonic Imaging*, vol. 24, no. 3, pp. 129–138, 2002.

[39] K. Nightingale, "Acoustic radiation force impulse (ARFI) imaging: A review," *Current medical imaging reviews*, vol. 7, no. 4, pp. 328–339, 2011.

[40] A. P. Sarvazyan, O. V. Rudenko, S. D. Swanson, J. B. Fowlkes, and S. Y. Emelianov, "Shear wave elasticity imaging: A new ultrasonic technology of medical diagnostics," *Ultrasound in Med. & Biol.*, vol. 24, no. 9, pp. 1419–1435, 1998.

[41] M. L. Palmeri, Y. Deng, N. C. Rouze, and K. R. Nightingale, "Dependence of shear wave spectral content on acoustic radiation force excitation duration and spatial beamwidth," *IEEE International Ultrasonics Symposium*, pp. 1105–1108, 2014.

[42] J.-L. Gennisson, T. Deffieux, M. Fink, and M. Tanter, "Ultrasound elastography: Principles and techniques," *Diagnostic and Interventional Imaging*, vol. 94, no. 5, pp. 487–495, 2013.

[43] R. M. Lerner, K. J. Parker, J. Holen, R. Gramiak, and R. C. Waag, "Sono-elasticity: Medical elasticity images derived from ultrasound signals in mechanically vibrated targets," in *Acoustical Imaging*, L. W. Kessler, Ed. Springer Boston MA, 1988, vol. 16, ISBN: 978-1-4612-8051-4.

[44] T. A. Krouskop, D. R. Dougherty, and F. S. Vinson, "A pulsed doppler ultrasonic system for making noninvasive measurements of the mechanical properties of soft tissue," *Journal of Rehabilitation Research and Development*, vol. 24, no. 2, pp. 1–8, 1987.

[45] K. J. Parker, S. R. Huang, R. Musulin, and R. M. Lerner, "Tissue response to mechanical vibrations for sonoelasticity imaging," *Ultrasound in Medicine & Biology*, vol. 16, no. 3, pp. 241–246, 1990.

[46] R. M. Lerner, S. R. Huang, and K. J. Parker, "Sonoelasticity images derived from ultrasound signals in mechanically vibrated tissues," *Ultrasound in Medicine & Biology*, vol. 16, no. 3, pp. 231–239, 1990.

[47] F. Lee, J. P. Bronson, R. M. Lerner, K. J. Parker, S. R. Huang, and D. J. Roachr, "Sonoelasticity imaging: Results in in vitro tissue specimens," *Radiology*, vol. 181, no. 1, pp. 237–239, 1991.

[48] Y. Yamakoshi, J. Sato, and T. Sato, "Ultrasonic imaging of internal vibration of soft tissue under forced vibration," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 37, no. 2, pp. 45–53, 1990.

[49] Z. Wu, L. S. Taylor, D. J. Rubens, and K. J. Parker, "Sonoelastographic imaging of interference patterns for estimation of the shear velocity of homogeneous biomaterials," *Physics in Medicine & Biology*, vol. 49, no. 6, pp. 911–922, 2004.

[50] K. Hoyt, B. Castaneda, and K. J. Parker, "Two-dimensional sonoelastographic shear velocity imaging," *Physics in Medicine & Biology*, vol. 34, no. 2, pp. 276–288, 2008.

[51] J. Ophir, I. Céspedes, H. Ponnekanti, Y. Yazdi, and X. Li, "Elastography: A quantitative method for imaging the elasticity of biological tissues," *Ultrasonic Imaging*, vol. 13, no. 2, pp. 111–134, 1991.

[52] J. Ophir, B. Garra, F. Kallel, *et al.*, "Elastographic imaging," *Ultrasound in Medicine & Biology*, vol. 26, S23–S29, 2000.

[53] C. F. Dietrich, R. G. Barr, A. Farrokh, *et al.*, "Strain elastography - how to do it," *Ultrasound International Open*, vol. 3, no. 4, pp. 137–149, 2017.

[54] R. S. Lazebnik, "Tissue strain analytics virtual touch tissue imaging and quantification," Siemens Medical Solutions, Mountain View, CA USA, White Paper, 2008.

[55] S. Catheline, F. Wu, and M. Fink, "A solution to diffraction biases in sonoelasticity: The acoustic impulse technique," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2941–2950, 1999.

[56] L. Sandrin, M. Tanter, J.-L. Gennisson, S. Catheline, and M. Fink, "Shear elasticity probe for soft tissues with 1-D transient elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 49, no. 4, pp. 436–446, 2002.

[57] L. Sandrin, M. Tanter, S. Catheline, and M. Fink, "Shear modulus imaging with 2-D transient elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 49, no. 4, pp. 426–435, 2002.

[58] L. Sandrin, B. F. amd Jean-Michel Hasquenoph, S. Yon, *et al.*, "Transient elastography: A new noninvasive method for assessment of hepatic fibrosis," *Ultrasound in Medicine & Biology*, vol. 29, no. 12, pp. 1705–1713, 2003.

[59] K. Nightingale, S. McAleavey, and G. Trahey, "Shear-wave generation using acoustic radiation force: In vivo and ex vivo results," *Ultrasound in Medicine & Biology*, vol. 29, no. 12, 2003.

[60] M. L. Palmeri, M. H. Wang, J. J. Dahl, K. D. Frinkley, and K. R. Nightingale, "Quantifying hepatic shear modulus in vivo using acoustic radiation force," *Ultrasound in Medicine & Biology*, vol. 34, no. 4, pp. 546–558, 2008.

[61] J. Bercoff, M. Tanter, and M. Fink, "Supersonic shear imaging: A new technique for soft tissue elasticity mapping," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 51, no. 4, pp. 396–409, 2004.

[62] M. Tanter, J. Bercoff, L. Sandrin, and M. Fink, "Ultrafast compound imaging for 2-D motion vector estimation: Application to transient elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 49, no. 10, pp. 1363–1374, 2002.

[63] M. Tanter, J. Bercoff, A. Athanasiou, *et al.*, "Quantitative assessment of breast lesion viscoelasticity: Initial clinical results using supersonic shear imaging," *Ultrasound in Medicine & Biology*, vol. 34, no. 9, pp. 1373–1386, 2008.

[64] M. Muller, J.-L. Gennisson, T. Deffieux, M. Tanter, and M. Fink, "Quantitative viscoelasticity mapping of human liver using supersonic shear imaging: Preliminary in vivo feasibility study," *Ultrasound in Medicine & Biology*, vol. 35, no. 2, pp. 219–229, 2009.

[65] L. Marais, M. Pernot, H. Khettab, *et al.*, "Arterial stiffness assessment by shear wave elastography and ultrafast pulse wave imaging: Comparison with reference techniques in normotensives and hypertensives," *Ultrasound in Medicine & Biology*, vol. 45, no. 3, pp. 758–772, 2019.

[66] S. McAleavey, M. Menon, and E. Elegbe, "Shear modulus imaging with spatially modulated ultrasound radiation force," *Ultrasonic Imaging*, vol. 31, no. 4, pp. 217–234, 2009.

[67] S. McAleavey, E. Collins, J. Kelly, E. Elegbe, and M. Menon, "Validation of SMURF estimation of shear modulus in hydrogels," *Ultrasonic Imaging*, vol. 31, no. 2, pp. 131–150, 2009.

[68] P. J. H. amd Stephen J Rosenzweig, K. R. Nightingale, and G. E. Trahey, "Single- and multiple-track-location shear wave and acoustic radiation force impulse imaging: Matched comparison of contrast, contrast-to-noise ratio and resolution," *Ultrasound in Medicine & Biology*, vol. 41, no. 4, pp. 1043–1057, 2015.

[69] M. Fatemi and J. F. Greenleaf, "Ultrasound-stimulated vibro-acoustic spectrography," *Science*, vol. 280, no. 5360, pp. 82–85, 1998.

[70] M. W. Urban, C. Chalek, R. R. Kinnick, *et al.*, "Implementation of vibro-acoustography on a clinical ultrasound system," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 58, no. 6, pp. 1169–1181, 1998.

[71] M. W. Urban, A. Alizad, W. Aquino, J. F. Greenleaf, and M. Fatemi, "A review of vibro-acoustography and its applications in medicine," *Current Medical Imaging Reviews*, vol. 7, no. 4, pp. 350–359, 2011.

[72] E. E. Konofagou and K. Hynynen, "Localized harmonic motion imaging: Theory, simulations and experiments," *Physics in Medicine & Biology*, vol. 29, no. 10, pp. 1405–1413, 2003.

[73] C. Maleke and E. E. Konofagou, "Harmonic motion imaging for focused ultrasound (HMIFU): A fully integrated technique for sonication and monitoring of thermal ablation in tissues," *Physics in Medicine & Biology*, vol. 53, no. 6, pp. 1773–1793, 2008.

[74] Y. Han, S. Wang, T. Payen, and E. Konofagou, "Fast lesion mapping during HIFU treatment using harmonic motion imaging guided focused ultrasound (HMIgFUS) in vitro and in vivo," *Physics in Medicine & Biology*, vol. 62, no. 8, pp. 3111–3123, 2017.

[75] E. E. Konofagou, C. Maleke, and J. Vappou, "Harmonic motion imaging (HMI) for tumor imaging and treatment monitoringe," *Current Medical Imaging Reviews*, vol. 8, no. 1, pp. 16–26, 2012.

[76] Z. Hah, C. Hazard, Y. T. Cho, D. Rubens, and K. Parker, "Crawling waves from radiation force excitation," *Ultrasonic Imaging*, vol. 32, no. 3, pp. 177–189, 2010.

[77] C. Hazard, Z. Hah, D. Rubens, and K. Parker, "Integration of crawling waves in an ultrasound imaging system. part 1: System and design considerations," *Ultrasound in Medicine & Biology*, vol. 38, no. 2, pp. 296–311, 2012.

[78] Z. Hah, C. Hazard, B. Millis, C. Barry, D. Rubens, and K. Parker, "Integration of crawling waves in an ultrasound imaging system. part 2: Signal processing and applications," *Ultrasound in Medicine & Biology*, vol. 38, no. 2, pp. 312–323, 2012.

[79] M. W. Urban, "Production of acoustic radiation force using ultrasound: Methods and applications," *Expert Review of Medical Devices*, vol. 15, no. 11, pp. 819–834, 2018.

[80] D. Cacko and M. Lewandowski, "Shear wave elastography implementation on a portable research ultrasound system: Initial results," *Applied Sciences*, vol. 12, no. 6210, 2022.

[81] P. Kijanka and M. W. Urban, "Local phase velocity based imaging of viscoelastic phantoms and tissues," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 3, pp. 389–405, 2021.

[82] G. Ferraioli, P. Parekh, A. B. Levitov, and C. Filice, "Shear wave elastography for evaluation of liver fibrosis," *Journal of Ultrasound in Medicine*, vol. 43, no. 4, pp. 800–807, 2018.

[83] R. G. Barr, "Shear wave liver elastography," *Abdominal Radiology*, vol. 43, no. 4, pp. 800–807, 2018.

[84] R. G. Barr, "Future of breast elastography," *Ultrasonography*, vol. 38, no. 2, pp. 93–105, 2019.

[85] D. J. Tyloch, J. F. Tyloch, J. Adamowicz, *et al.*, "Elastography in prostate gland imaging and prostate cancer detection," *Medical Ultrasonography*, vol. 20, no. 4, pp. 515–523, 2018.

[86] C.-K. Zhao and H.-X. Xu, "Ultrasound elastography of the thyroid: Principles and current status," *Ultrasonography*, vol. 38, no. 2, pp. 106–124, 2019.

[87] B. Zhou, X. Yang, X. Zhang, W. J. Curran, and T. Liu, "Ultrasound elastography for lung disease assessment," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 11, pp. 2249–2257, 2020.

[88] R. Iwasaki, R. Takagi, R. Nagaoka, *et al.*, "Monitoring of high-intensity focused ultrasound treatment by shear wave elastography induced by two-dimensional-array therapeutic transducer," *Japanese Journal of Applied Physics*, vol. 55, no. 7S1, 2016.

[89] D. Cosgrove, F. Piscaglia, J. Bamber, *et al.*, "EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. part 2: Clinical applications," *Ultraschall Med.*, vol. 34, no. 3, 2013.

[90] R. G.Barr, K. Nakashima, DominiqueAmy, *et al.*, "WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 2: Breast," *Ultrasound in Medicine & Biology*, vol. 41, no. 5, pp. 1148–1160, 2015.

[91] G. Ferraioli, C. Filice, L. Castera, *et al.*, "WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 3: Liver," *Ultrasound in Medicine & Biology*, vol. 41, no. 5, pp. 1161–1179, 2015.

[92] D. Cosgrove, R. Barr, J. Bojunga, *et al.*, "WFUMB guidelines and recommendations on the clinical use of ultrasound elastography: Part 4. thyroid," *Ultrasound in Medicine & Biology*, vol. 43, no. 1, pp. 4–26, 2017.

[93] R. G.Barr, D. Cosgrove, M. Brock, *et al.*, "WFUMB guidelines and recommendations on the clinical use of ultrasound elastography: Part 5. prostate," *Ultrasound in Medicine & Biology*, vol. 43, no. 1, pp. 27–48, 2017.

[94] B. P. Nelson and A. Sanghvi, "Point-of-care ultrasound," *European Journal of Trauma and Emergency Surgery*, vol. 42, no. 2, pp. 139–150, 2016.

[95] B. Shin, S. Jeon, J. Ryu, and H. J. Kwon, "Point-of-care ultrasound," *Elastography for portable ultrasound*, vol. 8, no. 1, pp. 101–116, 2018.

[96] Clarius. "Real-time elastography now available on clarius handheld ultrasound." (2019), [Online]. Available: `https://clarius.com/blog/real-time-elastography-now-available-on-clarius-handheld-ultrasound/` (visited on 08/05/2022).

[97] R. Zahiri, "Elastography," Clarius, Vancouver, Canada, White Paper, 2019.

[98] D. Cacko, M. Walczak, and M. Lewandowski, "Low-power ultrasound imaging on mixed FPGA/GPU systems," *IEEE 2018, IEEE Joint Conference - Acoustics*, 2018.

[99] B. Y. S. Yiu, M. Walczak, M. Lewandowski, and A. C. H. Yu, "Live ultrasound color-encoded speckle imaging platform for real-time complex flow visualization in vivo," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 4, pp. 656–668, 2019.

[100] M. Lewandowski, M. Walczak, P. Karwat, B. Witek, and P. Karłowicz, "Research and medical transcranial doppler system," *Archives of Acoustics*, vol. 41, no. 4, pp. 773–781, 2016.

[101] M. Lewandowski, M. Walczak, B. Witek, J. Rozbicki, and T. Steifer, "A GPU-based portable phased-array system with full-matrix capture," *IUS 2018, IEEE International Ultrasonics Symposium*, pp. 1–3, 2018.

[102] M. Lewandowski, M. Walczak, B. Witek, J. Rozbicki, and T. Steifer, "Demonstrator przenosnego systemu phased-array z funkcją full-matrix capture," *Badania Nieniszczące i Diagnostyka*, vol. 3, pp. 70–71, 2018.

[103] us4us. "Products." (2022), [Online]. Available: `https://us4us.eu/products/` (visited on 08/06/2022).

[104] Exo Imaging. "pMUT transducer - exo cello." (2022), [Online]. Available: `https://www.exo.inc/technology/` (visited on 08/06/2022).

[105] Y. Deng, M. L. Palmeri, N. C. Rouze, S. J. Rosenzweig, M. F. Abdelmalek, and K. R. Nightingale, "Analyzing the impact of increasing mechanical index and energy deposition on shear wave speed reconstruction in human liver," *Ultrasound in Medicine & Biology*, vol. 41, no. 7, pp. 1948–1957, 2015.

[106] M. L. Palmeri, S. A. McAleavey, G. E. Trahey, and K. R. Nightingale, "Ultrasonic tracking of acoustic radiation force-induced displacements in homogeneous media," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 53, no. 7, pp. 1300–1313, 2006.

[107] Y. Deng, N. C. Rouze, M. L. Palmeri, and K. R. Nightingale, "On system-dependent sources of uncertainty and bias in ultrasonic quantitative shear-wave imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 3, pp. 381–393, 2016.

[108] A. Ouared, E. Montagnon, S. Kazemirad, L. Gaboury, A. Robidoux, and G. Cloutier, "Frequency adaptation for enhanced radiation force amplitude in dynamic elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 62, no. 8, pp. 1300–1313, 2015.

[109] P. Song, H. Zhao, A. Manduca, M. W. Urban, J. F. Greenleaf, and S. Chen, "Comb-push ultrasound shear elastography (CUSE): A novel method for two-dimensional shear elasticity imaging of soft tissues," *IEEE Transactions on Medical Imaging*, vol. 31, no. 9, pp. 1821–1832, 2012.

[110] P. Song, M. W. Urban, A. Manduca, H. Zhao, J. F. Greenleaf, and S. Chen, "Comb-push ultrasound shear elastography (CUSE) with various ultrasound push beams," *IEEE Transactions on Medical Imaging*, vol. 32, no. 8, pp. 1435–1447, 2013.

[111] P. Song, A. Manduca, H. Zhao, M. W. Urban, J. F. Greenleaf, and S. Chen, "Fast shear compounding using robust 2-D shear wave speed calculation and multi-directional filtering," *Ultrasound in Medicine & Biology*, vol. 40, no. 6, pp. 1343–1355, 2014.

[112] A. Nabavizadeh, P. Song, S. Chen, J. F. Greenleaf, and M. W. Urban, "Multi-source and multi-directional shear wave generation with intersecting steered ultrasound push beams," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 62, no. 4, pp. 647–662, 2015.

[113] A. Nabavizadeh, J. F. Greenleaf, M. Fatemi, and M. W. Urban, "Optimized shear wave generation using hybrid beamforming methods," *Ultrasound in Medicine & Biology*, vol. 40, no. 1, pp. 188–199, 2014.

[114] F. Feng, S. Goswami, S. Khan, and S. A. McAleavey, "Shear wave elasticity imaging using nondiffractive bessel apodized acoustic radiation force," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 12, pp. 3528–3539, 2021.

[115] J.-L. Gennisson and M. Tanter, "Supersonic shear imaging," in *Ultrasound Elastography for Biomedical Applications and Medicine*, I. Nenadic, N. Urban, J. Greenleaf, J.-L. Gennisson, M. Bernal, and M. Tanter, Eds. Wiley & Sons, 2019, ISBN: 9781119021551.

[116] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 3, pp. 489–506, 2009.

[117] M. Capriotti, J. F. Greenleaf, and M. W. Urban, "Time-aligned plane wave compounding methods for high-frame-rate shear wave elastography: Experimental validation and performance assessment on tissue phantoms," *Ultrasound in Medicine & Biology*, vol. 47, no. 7, pp. 1931–1948, 2021.

[118] C. Papadacci, M. Pernot, M. Couade, M. Fink, and M. Tanter, "High-contrast ultrafast imaging of the heart," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 61, no. 2, pp. 288–301, 2014.

[119] J. R. Doherty, J. J. Dahl, and G. E. Trahey, "Harmonic tracking of acoustic radiation force induced displacements," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 60, no. 11, pp. 2347–2358, 2013.

[120] P. Song, H. Zhao, M. W. Urban, *et al.*, "Improved shear wave motion detection using pulse-inversion harmonic imaging with a phased array transducer," *IEEE Transactions on Medical Imaging*, vol. 32, no. 12, pp. 2299–2310, 2013.

[121] P. Song, M. W. Urban, A. Manduca, J. F. Greenleaf, and S. Chen, "Coded excitation plane wave imaging for shear wave motion detection," *IEEE Transactions on Medical Imaging*, vol. 62, no. 7, pp. 1356–1372, 2015.

[122] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. A. Jensen, and O. Bernard, "Plane-wave imaging challenge in medical ultrasound," *Proceedings of 2016 IEEE International Ultrasonics Symposium (IUS)*, 2016.

[123] Y. Deng, N. C. Rouze, M. L. Palmeri, and K. R. Nightingale, "Ultrasonic shear wave elasticity imaging sequencing and data processing using a verasonics research scanner," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 1, pp. 164–176, 2017.

[124] C. Kasai, K. Namekawa, A. Koyano, and R. Omoto, "Real-time two-dimensional blood flow imaging using an autocorrelation technique," *IEEE Transactions on Sonics and Ultrasonics*, vol. 32, no. 3, pp. 458–464, 1985.

[125] T. Loupas, R. B. Peterson, and R. W. Gill, "Experimental evaluation of velocity and power estimation for ultrasound blood flow imaging, by means of a two-dimensional autocorrelation approach," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 42, no. 4, pp. 689–699, 1995.

[126] G. F. Pinton, J. J. Dahl, and G. E. Trahey, "Rapid tracking of small displacements with ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 53, no. 6, pp. 1103–1117, 2006.

[127] H.-K. Lee, J. F. Greenleaf, and M. W. Urban, "A new plane wave compounding scheme using phase compensation for motion detection," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 2, pp. 702–710, 2022.

[128] F. Viola and W. F. Walker, "A comparison of the performance of time-delay estimators in medical ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 50, no. 4, pp. 392–401, 2003.

[129] B. Byram, G. E. Trahey, and M. Palmeri, "Bayesian speckle tracking. part II: Biased ultrasound displacement estimation," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 60, no. 1, pp. 144–157, 2013.

[130] D. M. Dumont and B. C. Byram, "Robust tracking of small displacements with a bayesian estimator," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 1, pp. 20–34, 2016.

[131] M. D. Horeh, A. Asif, and H. Rivaz, "Analytical minimization-based regularized subpixel shear-wave tracking for ultrasound elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 2, pp. 285–296, 2018.

[132] D. C. Mellema, P. Song, R. R. Kinnick, *et al.*, "Probe oscillation shear wave elastography: Initial in vivo results in liver," *IEEE Transactions on Medical Imaging*, vol. 37, no. 5, pp. 1214–1223, 2018.

[133] P. Song, "Innovations in ultrasound shear wave elastography," Ph.D. dissertation, Mayo Clinic College of Medicine, 2014.

[134] A. Manduca, D. S. Lake, S. A. Kruse, and R. L. Ehman, "Spatio-temporal directional filtering for improved inversion of mr elastography images," *Medical Image Analysis*, vol. 7, no. 4, pp. 465–473, 2003.

[135] T. Deffieux, J.-L. Gennisson, J. Bercoff, and M. Tanter, "On the effects of reflected waves in transient shear wave elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 58, no. 10, pp. 2032–2035, 2011.

[136] J.-L. Gennisson, J. Provost, T. Deffieux, *et al.*, "4-D ultrafast shear-wave imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 62, no. 6, pp. 1059–1065, 2015.

[137] S. L. Lipman, N. C. Rouze, M. L. Palmeri, and K. R. Nightingale, "Evaluating the improvement in shear wave speed image quality using multidimensional directional filters in the presence of reflection artifacts," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 8, pp. 1049–1063, 2016.

[138] M. H. Wang, M. L. Palmeri, V. M. Rotemberg, N. C. Rouze, and K. R. Nightingale, "Improving the robustness of time-of-flight based shear wave speed reconstruction methods using RANSAC in human liver in vivo," *Ultrasound in Medicine & Biology*, vol. 36, no. 5, pp. 802–813, 2010.

[139] N. C. Rouze, M. H. Wang, and M. L. P. andand Kathryn R. Nightingale, "Parameters affecting the resolution and accuracy of 2D quantitative shear wave images," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 59, no. 8, pp. 1729–1740, 2012.

[140] J. Racedo and M. W. Urban, "Evaluation of reconstruction parameters for 2-D comb-push ultrasound shear wave elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 2, pp. 254–263, 2019.

[141] J. McLaughlin and D. Renzi, "Shear wave speed recovery in transient elastography and supersonic imaging using propagating fronts," *Inverse Problems*, vol. 22, no. 2, pp. 681–706, 2006.

[142] H.-K. Lee, D. Kong, K. Choi, R. Mislati, and M. M. Doyley, "A robust and fast method for 2-D shear wave speed calculation," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 7, pp. 2351–2360, 2021.

[143] N. C. Rouze, M. H. Wang, M. L. Palmeri, and K. R. Nightingale, "Robust estimation of time-of-flight shear wave speed using a radon sum transformation," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, no. 12, pp. 2662–2670, 2010.

[144] C. A. Carrascal, S. Chen, A. Manduca, J. F. Greenleaf, and M. W. Urban, "Improved shear wave group velocity estimation method based on spatiotemporal peak and thresholding motion search," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 4, pp. 660–668, 2017.

[145] Y. Li, Q. Lv, J. Dai, Y. Tian, and J. Guo, "Shear wave velocity estimation using the real-time curve tracing method in ultrasound elastography," *Applied Sciences*, vol. 11, no. 5, 2021.

[146] U.S. Food and Drug Administration, *Marketing clearance of diagnostic ultrasound systems and transducers: Guidance for industry and food and drug administration staff*, 2019.

[147] G. R. Harris, "Hydrophone measurements in diagnostic ultrasound fields," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 35, no. 2, pp. 87–101, 1988.

[148] G. R. Harris, S. M. Howard, A. M. Hurell, *et al.*, "Hydrophone measurements for biomedical ultrasound applications: A review," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 70, no. 2, pp. 85–101, 2023.

[149] M. L. Palmeri and K. R. Nightingale, "Image quality, tissue heating, and frame rate trade-offs in acoustic radiation force impulse imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 51, no. 5, pp. 551–565, 2004.

[150] B. J. Fahey, M. L. Palmeri, and G. E. Trahey, "Frame rate considerations for real-time abdominal acoustic radiation force impulse imaging," *Ultrasonic Imaging*, vol. 28, no. 4, pp. 193–210, 2006.

[151] R. R. Bouchard, J. J. Dahl, S. J. Hsu, M. L. Palmeri, and G. E. Trahey, "Image quality, tissue heating, and frame rate trade-offs in acoustic radiation force impulse imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 1, pp. 63–76, 2009.

[152] S. A. McAleavey, L. O. Osapoetra, and J. Langdon, "Shear wave arrival time estimates correlate with local speckle pattern," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 62, no. 12, pp. 2054–2067, 2015.

[153] P. J. Hollender, S. J. Rosenzweig, K. R. Nightingale, and G. E. Trahey, "Single- and multiple- track location shear wave and acoustic radiation force impulse imaging: Matched comparison of contrast, CNR, and resolution," *Ultrasound in Medicine & Biology*, vol. 41, no. 4, pp. 1043–1057, 2015.

[154] S. A. McAleavey, "Single tracking location shear wave elastography," in *Ultrasound Elastography for Biomedical Applications and Medicine*, I. Nenadic, N. Urban, J. Greenleaf, J.-L. Gennisson, M. Bernal, and M. Tanter, Eds. Wiley & Sons, 2019, ISBN: 9781119021551.

[155] B. R. Chintada, R. Rau, and O. Goksel, "Phase-aberration correction in shear-wave elastography imaging using local speed-of-sound adaptive beamforming," *Frontiers in Physics*, vol. 9, no. 690385, 2021.

[156] H. Zhao, P. Song, M. W. Urban, *et al.*, "Bias observed in time-of-flight shear wave speed measurements using radiation force of a focused ultrasound beam," *Ultrasound in Medicine & Biology*, vol. 37, no. 11, pp. 1884–1892, 2012.

[157] M. Walczak, M. Lewandowski, and N. Żołek, "Optimization of real-time ultrasound PCIe data streaming and OpenCL processing for SAFT imaging," *Proceedings of 2013 IEEE International Ultrasonics Symposium (IUS)*, 2013.

[158] B. Y. S. Yiu, I. K. H. Tsang, and A. C. H. Yu, "GPU-based beamformer: Fast realization of plane wave compounding and synthetic aperture imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 58, no. 6, pp. 1698–1705, 2011.

[159] B. Y. S. Yiu and A. C. H. Yu, "GPU-based minimum variance beamformer for synthetic aperture imaging of the eye," *Ultrasound in Medicine & Biology*, vol. 41, no. 3, pp. 871–883, 2015.

[160] D. Hyun, Y. L. Li, I. Steinberg, M. Jakovljevic, T. Klap, and J. J. Dahl, "An open source GPU-based beamformer for real-time ultrasound imaging and applications," *2019 IEEE International Ultrasonics Symposium (IUS)*, 2019.

[161] M. Lewandowski, P. Jarosik, Y. Tasinkevych, and M. Walczak, "Efficient GPU implementation of 3D spectral domain synthetic aperture imaging," *2020 IEEE International Ultrasonics Symposium (IUS)*, 2020.

[162] L. W. Chang, K. H. Hsu, and P. C. Li, "Graphics processing unit-based high-frame-rate color doppler ultrasound processing," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 9, pp. 1856–1860, 2009.

[163] A. J. Y. Chee, B. Y. S. Yiu, and A. C. H. Yu, "A GPU-parallelized eigen-based clutter filter framework for ultrasound color flow imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 1, pp. 150–163, 2017.

[164] P. Jarosik and M. Lewandowski, "The feasibility of deep learning algorithms integration on a GPU-based ultrasound research scanner," *2017 IEEE International Ultrasonics Symposium (IUS)*, 2017.

[165] S. Rosenzweig, M. Palmeri, and K. Nightingale, "GPU-based real-time small displacement estimation with ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 58, no. 2, pp. 399–405, 2011.

[166] M. Lewandowski, "Medical ultrasound digital signal processing in the GPU computing era," in *Computer Vision in Medical Imaging*, C. H. Chen, Ed. World Scientific Publishing Company, 2013, ISBN: 978-9814460934.

[167] E. Alcaín, P. R. Fernández, R. Nieto, *et al.*, "Hardware architectures for real-time medical imaging," *Electronics*, vol. 24, no. 10, 2021.

[168] B. Witek, M. Walczak, and M. Lewandowski, "Characterization of the STHV748 integrated pulser for generating push sequences," *Proceedings of 2015 IEEE International Ultrasonics Symposium (IUS)*, 2015.

[169] STMicroelectronics, *16 channel ±100 V, ±2/±4 A, 5/3 level with RTZ, T/R switch, high-speed ultrasound pulser with integrated transmit beamformer*, Data brief, 2020.

[170] Analog Devices, *LTM4622: Dual Ultrathin 2.5A or Single 5A Step-Down DC/DC µModule Regulator*, Datasheet, 2020.

[171] Analog Devices, *LTM8049: Dual SEPIC or Inverting µModule DC/DC Converter*, Datasheet, 2018.

[172] Analog Devices, *ADP1754/ADP1755: 1.2 A, Low VIN, Low Dropout Linear Regulator*, Datasheet, 2014.

[173] Analog Devices, *LT3015 series: 1.5A, Low Noise, Negative Linear Regulator with Precision Current Limit*, Datasheet, 2014.

[174] Skyworks Solutions, Inc., *Si5338: I2C-Programmable Any-Frequency, Any-Output Quad Clock Generator*, 2021.

[175] Skyworks Solutions, Inc., *Si5330: 1.8/2.5/3.3 V Low-Jitter, Low-Skew Clock Buffer/Level Translator*, 2021.

[176] Skyworks Solutions, Inc., *Si53340-45 Data Sheet*, 2021.

[177] Cypress Semiconductor, *3.0 V/1.8 V, 64 Mb (8 MB)/128 Mb (16 MB), HyperRAM™ Self-Refresh DRAM*, Datasheet, 2018.

[178] Texas Instruments, *TMP75C 1.8-V Digital Temperature Sensor with Two-Wire Interface and Alert*, Datasheet, 2014.

[179] J. A. Flynn, P. Kaczkowski, K. Linkhart, and R. E. Daigle, "Arbitrary waveforms using a tri-state transmit pulser," *2013 IEEE International Ultrasonics Symposium (IUS)*, 2013.

[180] Verasonics Inc., *Arbitrary Waveform Generation with the Verasonics Research Ultrasound Platform*, White Paper, 2016.

[181] G. Ferraioli, C. Tinelli, Mabel, *et al.*, "Reproducibility of real-time shear wave elastography in the evaluation of liver elasticity," *European Journal of Radiology*, vol. 81, no. 11, pp. 3102–3106, 2012.

[182] J.-M. Correas, A.-M. Tissier, A. Khairoune, *et al.*, "Prostate cancer: Diagnostic performance of real-time shear-wave elastography," *Radiology*, vol. 275, no. 1, 2014.

[183] B. Procopet, A. Berzigotti, J. G. Abraldes, *et al.*, "Real-time shear-wave elastography: Applicability, reliability and accuracy for clinically significant portal hypertension," *Journal of Hepatology*, vol. 62, no. 5, pp. 1068–1075, 2015.

[184] Intel Corporation. "Export compliance metrics for Intel microprocessors." (2022), [Online]. Available: `https://www.intel.com/content/www/us/en/support/articles/000005755/processors.html` (visited on 11/08/2022).

[185] D. B. Kirk and W.-m. W. Hwu, *Programming Massively Parallel Processors: A Hands-on Approach*. Elsevier, 2010, ISBN: 978-0-12-381472-2.

[186] J. Nickolls and W. J. Dally, "The GPU computing era," *IEEE Micro*, vol. 30, no. 2, pp. 56–69, 2010.

[187] NVidia, *CUDA C++ Programming Guide*, Design Guide, 2022.

[188] J. Sanders and E. Kandrot, *CUDA by example: an introduction to general-purpose GPU programming*. NVIDIA Corporation, 2010, ISBN: 978-0-13-138768-3.

[189] NVIDIA Corporation. "Cuda toolkit." (2022), [Online]. Available: `https://developer.nvidia.com/cuda-toolkit` (visited on 11/11/2022).

[190] "Numpy webpage." (2022), [Online]. Available: `https://numpy.org/` (visited on 11/11/2022).

[191] "Scipy webpage." (2022), [Online]. Available: `https://scipy.org/` (visited on 11/11/2022).

[192] P. A. Entschev. "Single-GPU CuPy speedups." (2019), [Online]. Available: `https://medium.com/rapids-ai/single-gpu-cupy-speedups-ea99cbbb0cbb` (visited on 11/11/2022).

[193] "Cupy webpage." (2022), [Online]. Available: `https://cupy.dev/` (visited on 11/11/2022).

[194] D. Cacko, P. Jarosik, and M. Lewandowski, "Real-time shear wave elastography implementation on a portable research ultrasound system with GPU-accelerated processing," *Proceedings of 2013 IEEE International Ultrasonics Symposium (IUS)*, 2023.

[195] M. Lewandowski, "Hardware-software partitioning of digital signal processing in ultrasound medical devices: A case study," in *Ultrasound Imaging*. Intech Open, 2011, ISBN: 978-9814460934.

[196] R. Kishimoto, M. Suga, M. Usumura, *et al.*, "Shear wave speed measurement bias in a viscoelastic phantom across six ultrasound elastography systems: A comparative study with transient elastography and magnetic resonance elastography," *Journal of Medical Ultrasonics*, vol. 49, no. 2, pp. 143–152, 2022.

[197] C. Edwards, E. Cavanagh, S. Kumar, V. Clifton, and D. Fontanarosa, "Intra-system reliability assessment of 2-dimensional shear wave elastography," *Applied Sciences*, vol. 11, no. 7, 2021.

[198] J. A. Hwang, W. K. Jeong, K. D. Song, K. A. Kang, and H. K. Lim, "2-D shear wave elastography for focal lesions in liver phantoms: Effects of background stiffness, depth and size of focal lesions on stiffness measurement," *Ultrasound in Medicine & Biology*, vol. 45, no. 12, 2019.

[199] E. Harris, R. Sinnatamby, E. O'Flynn, A. M. Kirby, and J. C. Bamber, "A cross-machine comparison of shear-wave speed measurements using 2D shear-wave elastography in the normal female breast," *Applied Sciences*, vol. 11, no. 20, 2021.

[200] P. Song, M. C. Macdonald, R. H. Behler, *et al.*, "Shear wave elastography on the GE LOGIQ E9 with comb-push ultrasound shear elastography (CUSE) and time aligned sequential tracking (TAST)," *Proceedings of 2014 IEEE International Ultrasonics Symposium*, 2014.

[201] G. Healthcare, "2d shear wave elastography logiq e9/e10/e10s," General Electric Company, White Paper, 2020.

[202] S. Imagine, "Aixplorer ultimate system specifications," SuperSonic Imagine S.A., 2017.

[203] Many authors, *Ultrasound Elastography for Biomedical Applications and Medicine*, I. Nenadic, N. Urban, J. Greenleaf, J.-L. Gennisson, M. Bernal, and M. Tanter, Eds. Wiley & Sons, 2019, ISBN: 9781119021551.

[204] J. Dahmani, C. Laporte, D. Pereira, P. Belanger, and Y. Petit, "Predictive model for designing soft-tissue mimicking ultrasound phantoms with adjustable elasticity," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 4, pp. 715–726, 2020.

[205] P. Song, M. Macdonald, R. Behler, *et al.*, "Two-dimensional shear-wave elastography on conventional ultrasound scanners with time-aligned sequential tracking (TAST) and comb-push ultrasound shear elastography (CUSE)," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 62, no. 2, pp. 290–302, 2015.

[206] P. Kijanka and M. W. Urban, "Local phase velocity based imaging: A new technique used for ultrasound shear wave elastography," *IEEE Transactions on Medical Imaging*, vol. 38, no. 4, pp. 894–908, 2019.

[207] Computerized Imaging Reference Systems, Inc., *Elasticity QA phantom models 049 & 049a*, 2011.

Appendix A: Phantom data

This appendix contains information about the phantom used in this thesis.

Overview

Experiments were carried out using elasticity QA phantom (Fig. 1) model 049A by CIRS (Norfolk, USA). The phantom contains stepped cylinder mass targets of known stiffness relative to the background material. Targets range in stiffness, diameter, and depth. There are targets of 4 various nominal stiffnesses, 6 diameters, and positioned at two depths, giving 48 different combinations in total. See the specifications below for details.

Figure 1: The QA elasticity phantom in a test setup with probe.

Specifications

The QA phantom model 049 by CIRS specifications are listed in the Table 1. Phantom drawings with dimensions are shown in Fig. 2

Table 1: CIRS 049A phantom specifications.

Parameter	Value
Dimensions	20 cm x 15 cm x 10 cm
Phantom weight	3 kg
Housing material	ABS Plastic
Membrane	Saran laminate
Scannable surface area	17 x 10 cm
Tissue mimicking material	Zerdine solid elastic hydrogel [1]
	Speed of Sound: 1540 – 1570 m/s
Young's modulus[2]	Background: 16.7 kPa
	Lesion Type I: 4.7 kPa
	Lesion Type II: 8.6 kPa
	Lesion Type III: 34.0 kPa
	Lesion Type IV: 70.9 kPa

[1] More information about the material can be found on the manufacturer's website.

[2] Nominal values declared by manufacturer.
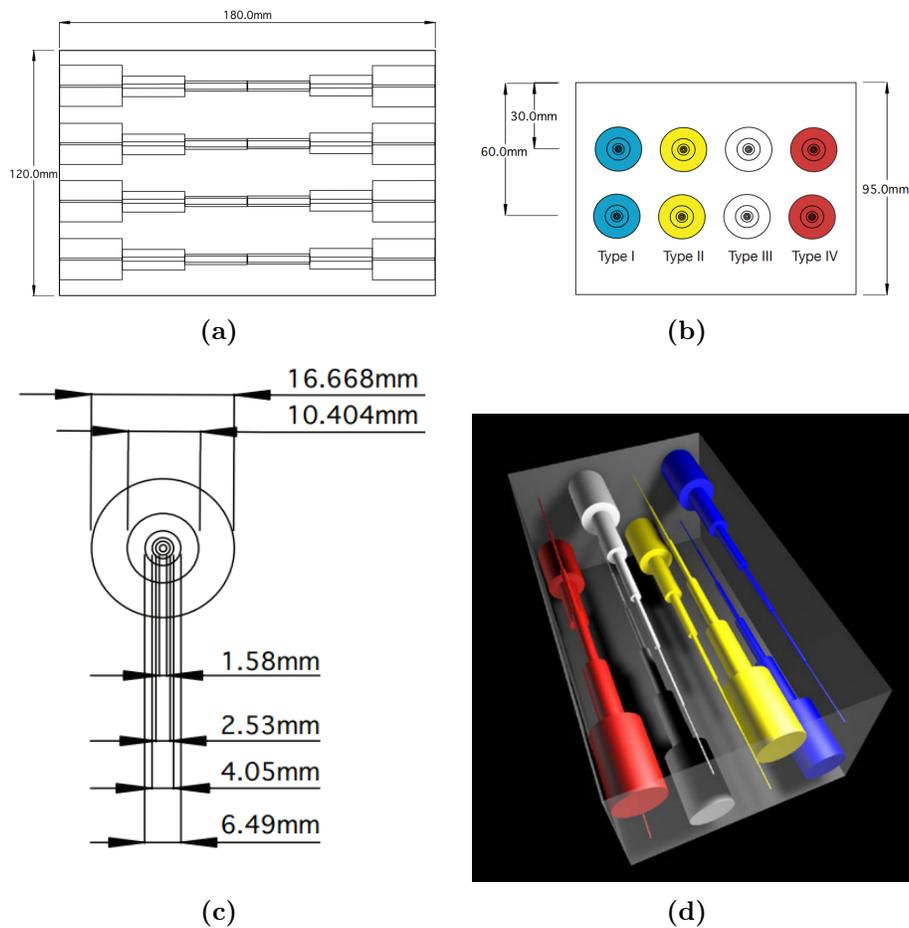
Figure 2: CIRS 049A phantom: (a) Top view. (b) Side view. (c) Target diameters. (d) Phantom visualization. Source [207].