Technical Report

TR-06-47

Inspection of copper canisters for spent nuclear fuel by means of ultrasound

Ultrasonic imaging, FSW monitoring with acoustic emission

Tadeusz Stepinski (editor), Tomas Olofsson,
Erik Wennerström
Uppsala University, Signals and Systems,
Department of Technical Sciences, Sweden

December 2006

Svensk Kärnbränslehantering AB
Swedish Nuclear Fuel
and Waste Management Co
Box 5864
SE-102 40 Stockholm Sweden
Tel 08-459 84 00
+46 8 459 84 00
Fax 08-661 57 19
+46 8 661 57 19

Inspection of copper canisters for spent nuclear fuel by means of ultrasound

Ultrasonic imaging, FSW monitoring with acoustic emission

Tadeusz Stepinski (editor), Tomas Olofsson,
Erik Wennerström
Uppsala University, Signals and Systems,
Department of Technical Sciences, Sweden

December 2006

Abstract

This report contains the research results concerning advanced ultrasound for the inspection of copper canisters for spent nuclear fuel obtained at Signals and Systems, Uppsala University in years 2005/2006.

In the first part of the report we propose a concept of monitoring of the friction stir welding (FSW) process by means of acoustic emission (AE) technique. First, we introduce the AE technique and then we present the principle of the system for monitoring the FSW process in cylindrical symmetry specific for the SKB canisters. We propose an omnidirectional circular array of ultrasonic transducers for receiving the AE signals generated by the FSW tool and the releases of the residual stress at canister's circumference. Finally, we review the theory of uniform circular arrays.

The second part of the report is concerned with synthetic aperture focusing technique (SAFT) characterized by enhanced spatial resolution. We evaluate three different approaches to perform imaging with less computational cost than that of the extended SAFT (ESAFT) method proposed in our previous reports. First, a sparse version of ESAFT is presented, which solves the reconstruction problem only for a small set of the most probable scatterers in the image. A frequency domain the ω-k SAFT algorithm, which relies on the far-field approximation is presented in the second part. Finally, a detailed analysis of the most computationally intense step in the ESAFT and the sparse 2D deconvolution is presented.

In the final part of the report we introduce basics of the 3D ultrasonic imaging that has a great potential in the inspection of the FSW welds. We discuss in some detail the three interrelated steps involved in the 3D ultrasonic imaging: data acquisition, 3D reconstruction, and 3D visualization.

Contents

Chapter 1

Introduction

by Tadeusz Stepinski

In this report we are presenting our recent results concerning monitoring of the friction stir welding (FSW) process used for sealing of copper canisters for spent nuclear fuel as well as the inspection of the FSW weld by means of ultrasound.

Our research activity in this project presented in this report was split into three separate tasks:

- FSW process monitoring using acoustic emission,
- High resolution synthetic aperture imaging,
- 3D ultrasonic imaging.

The first task, concerned with monitoring of the FSW process used for sealing the SKB canisters is reported in Chapter 2. In this chapter we propose a concept of monitoring the FSW process by means of acoustic emission (AE) technique. First, we introduce the AE technique and we present the principle of the system proposed for monitoring the FSW process in cylindrical symmetry specific for the SKB canisters. We propose an omnidirectional array of ultrasonic transducers for receiving the AE signals generated by the FSW tool and the releases of residual stress at canister's circumference. Theory of uniform circular arrays (UCAs) that are good candidates for this kind of application is reviewed in the second part of this chapter.

The second task, which is a continuation of our experimental and theoretical research concerning synthetic aperture focusing technique (SAFT) characterized by enhanced spatial resolution, is reported in Chapter 3. In this chapter we evaluate three different approaches to perform imaging with less computational cost than of the extended SAFT (ESAFT) method proposed in our previous reports. First, a sparse version of ESAFT is presented, which is based on the same model as ESAFT but it solves the reconstruction problem only for a small set of the most probable scatterers in the image. A frequency domain the ω-k SAFT algorithm, which relies on the far-field approximation is presented in the second part. Finally, an analysis of one of the most computationally intense steps in the ESAFT and sparse 2D deconvolution is presented.

In Chapter 4 we introduce basics of the 3D ultrasonic imaging that has a great potential in the inspection of friction stirred welds where a variety of flaws can be encountered. Unlike traditional 2D imaging, where cross-sections of the inspected volume are displayed and where the inspector must mentally reconstruct the involved volumetric structures, the 3D images have the advantage of being seen from whatever angle the inspector may desire. We discuss in some detail the three interrelated steps involved in 3D ultrasonic imaging: data acquisition, 3D reconstruction, and 3D visualization. When sufficient amount of data has been gathered the inspected volume can be reconstructed and the user can choose an appropriate visualization mode to view the data.

Chapter 2

Monitoring Friction Stir Welding with Acoustic Emission

by Tadeusz Stepinski

2.1 Introduction

The aim of this chapter is to present our concept of an automatic system for monitoring the friction stir welding (FSW) process applied to copper canisters.

This chapter consists of two separate parts:

- Brief presentation of acoustic emission (AE) technique,
- Theory of uniform circular arrays.

In the first part we introduce the AE technique focusing on sources of AE events and their sensing using piezoelectric transducers. We also present the principle of the system proposed for monitoring the FSW process in cylindrical symmetry specific for the SKB canisters. We propose an omnidirectional array of ultrasonic transducers for receiving the AE signals generated by the FSW tool and the releases of residual stress at canister's circumference. A transducer array, suitable for this task, should be characterized by good angular selectivity to separate the AE sources located at different points of canister circumference. Uniform circular arrays (UCAs) are good candidates for this kind of application. UCAs have applications in mobile telephony and they have a very rich literature concerning their applications for beamforming and direction of arrival estimation. We address those issues in the second part of this chapter.

2.2 Introduction to Acoustic Emission

Acoustic emission (AE) refers to the phenomenon of transient elastic wave generation due to a rapid release of strain energy caused by a structural or microstructural event in a solid material [1]. When deformation processes occur in materials, strain energy is released in the form of vibrations or stress waves. The basic mode of operation of a stress wave detection system is similar to that of passive sonar. When stress waves reach the specimen surface the small displacements they produce are detected by a number of ultrasonic transducers. The transducers that are coupled to the surface of the inspected structure collect the stress-wave, which has propagated from its source through the structure. The amplified signals from the transducers are then conditioned, preliminary analyzed and recorded. AE techniques enable detecting stress waves emitted from tiny energy sources, such as, crack propagation, dislocation motion or fiber disbonding in composites. A typical setup, used mainly in the laboratory AE applications, is illustrated in Fig. 2.1. Multiple sensors and amplifier channels that make possible localization of AE sources are used in most applications.

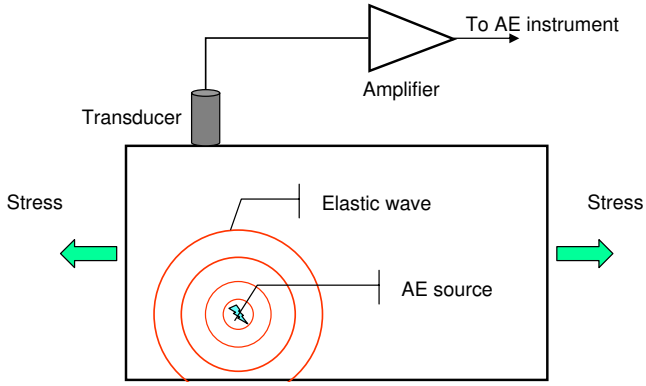


Figure 2.1: AE principle – an elastic wave emitted by a single event in the inspected specimen is received by the AE transducer.

AE inspection is unique in the sense that it relies on the signal that originates from the discontinuity itself. Most other NDE methods provide local excitation, some form of energy, to a structure and information about the structure is subsequently gained from the local response to the excitation. AE techniques detect the response of the structure to an external load (e.g., mechanical or thermal) whose purpose is to create stress throughout the structure, causing significant discontinuities, wherever they are, to emit acoustically. Because of this difference, acoustic emission is able to monitor large areas of a structure with a relatively small number of sensors. The technique can also provide real-time monitoring under normal operating conditions, [1].

2.2.1 AE sensors and signals

A wide range of basic transduction mechanisms can be used to achieve a sensor's functions – the detection of surface motion and the subsequent generation of an electrical signal. Capacitive transducers have been successfully used as acoustic emission sensors for special laboratory tests. Magnetostrictive transducers are preferred in some industrial applications.

However, almost all sensors applied for the reception of AE signals use piezoelectric elements for transduction, [1]. A piezoelectric ceramic disc, contained in a metal can, is bonded to the material's surface and produces an electrical signal when mechanically excited by the AE wave motions. The sharp leading edge of the wave excites the disc oscillations at its resonance frequencies to give a decaying transient signal.

Primarily, element thickness controls the frequencies at which the acoustic emission sensor has the highest electrical output, i.e. highest sensitivity, for a given input surface velocity. The predominant frequencies are chosen by selecting the dimensions of the disc and unwanted ones can be eliminated by using a narrow frequency band preamplifier. The predominant frequency is usually chosen to be between 50 and 500 kHz, since the background noise (mechanical, flow, etc.) is large below this band and attenuation of the acoustic surface wave increases at higher frequencies.

The other principal characteristic of an acoustic emission sensor is the active element diameter. The element diameter defines the area over which the sensor averages surface motion. Sensors have been designed with element diameters as small as 1 mm. However, larger diameters are more common. For waves resulting in uniform motion under the transducer, as is the case for a bulk longitudinal wave propagating in a direction perpendicular to the surface, increase of the

sensor element diameter results in the increased sensitivity. However, for waves traveling along the surface, e.g. Rayleigh waves, the element diameter strongly influences the sensor sensitivity as a function of wave frequency.

The high frequency output signal from an AE sensor is seldom recorded in its original form. In most applications it is fed into a threshold-level detector, which will only pass signals above the threshold level. For the idealized signal in the upper part of Fig. 2.2, the threshold level is shown, so the detector will only detect those sine waves in the pulse with the amplitudes above the threshold resulting in a binary signal shown in the lower part of Fig. 2.2. If the signal is then fed into a counter, a number of counts of will be recorded depending on the amplitude of the original AE response at the transducer output. Thus larger bursts will record larger counts.

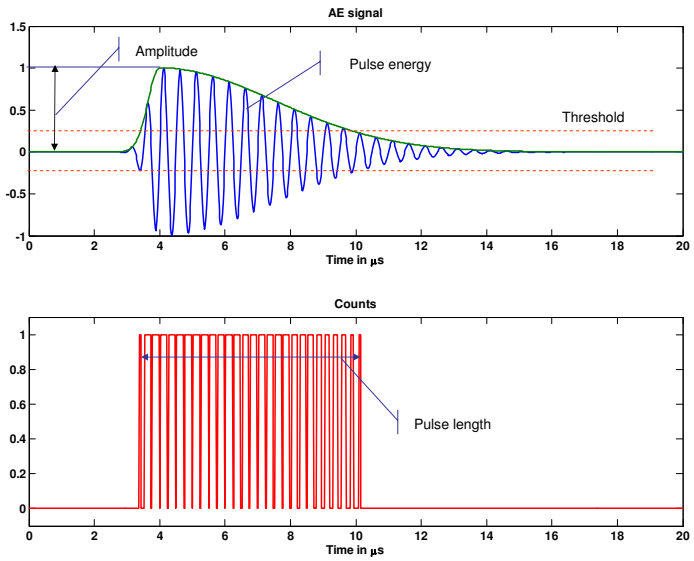


Figure 2.2: Typical response of an AE transducer. The AE response and its parameters (upper part) and the AE response after thresholding (lower part).

This method of counting, using the ringing of the piezoelectric transducer to give a measure of the pulse signal size is called ringdown counting and is most widely used in AE detection systems. However, other quantifiable features of AE signals can be used depending on the application at hand, for example:

- Rise time: the time taken to reach peak amplitude from the first present threshold voltage crossing of the signal;
- Peak amplitude: this can be related to the intensity of the source in the material producing an AE signal;
- RMS voltage: a measure of signal intensity.

Various advanced signal processing methods, such as, time series analysis, fast Fourier transform (FFT), and wavelet transform, have been applied to analysis AE signals, with the aim to extract the features of AE signals, [2, 3].

2.2.2 Sources of acoustic emission

The primary sources of acoustic emission in metals are the processes of plastic deformation and crack growth, which are discrete energy release mechanisms on a crystalline microstructure scale.

These micro-displacements produce bursts of stress-waves (acoustic emission) with frequency ranging from audio to high ultrasonic depending on the size and velocity of the source event.

Researchers in the field of acoustic emission agree that emission signals fall into two general classes, continuous and burst emission, [4]. In *continuous emission*, the stress-wave bursts with very low amplitude are observed that are unresolvable. High sensitivity is thus needed to detect them. Plastic deformation in a material, for instance, the ductile failure of flaw-free metals, produces this kind of low-level signal. *Burst emission* is a form of emission of much higher amplitude in which the individual stress-wave bursts can be distinguished. It occurs when sources of higher energy are operating. Crack growth is an important example of such. In metals, the concentration of stress around the tip of a crack produces a small region of plastic deformation known as the plastic zone. As this region expands under increasing stress, stress waves are generated.

In certain materials both the low and high energy sources may operate simultaneously. This is most likely when the material is non-homogeneous (multi-phase material), e.g., steels containing brittle carbides. Fibre-reinforced materials are classic examples of such complex materials. Fibre fracture is a high-energy source whereas matrix deformation and fibre disbonding are probably of low energy.

An important factor that has to be considered is the maximum previous stress level experienced by entire specimen. In metals it has been consistently observed that acoustic emission occurs only when this stress level is exceeded (*Kaiser effect*) [1].

AE due to crack growth

Cyclic stressing combined with regions of high stress concentration, create conditions for stable crack growth. The sudden separation of the fracture faces due to strain relief results in relatively large amplitude stress-wave emission that can be recorded as an AE event. The acoustic emission source mechanisms that exist at and close to the crack tip fall into the following two categories [5]:

Primary events that are directly associated with crack growth. They originate within the plastic zone, which extends a short distance ahead of the crack tip.

Secondary events that originate primarily from friction at points of contact between the clean fracture faces near the crack tip or from disbonding and fracture of corrosion products some way back from the crack tip.

The intensity of acoustic emission from crack growth depends on the magnitude of the accompanying micro-fracture events, which is a complicated function of crack size, material properties and applied stress. The acoustic emission is a measure of the rate of crack growth in time from which crack size may be inferred only in very special laboratory conditions.

AE emitted by cutting process

Research has shown that AE, due to its ability to detect stress waves generated by the sudden release of energy in deforming materials, is a very suitable tool for monitoring cutting processes. AE has been successfully used in laboratory tests to detect tool wear and fracture in single point turning operations [6].

Among the possible sources of AE during metal cutting processes we can name the following: plastic deformation during the cutting process in the workpiece and chip, frictional contact between the tool and the workpiece resulting in tool wear, collisions between chip and tool, chip breakage and, last but not least, tool fracture, [7, 8].

Based on the analysis of AE signal sources, AE derived from metal turning consists of continuous and transient signals, which have distinctly different characteristics. Continuous signals are associated with shearing in the primary zone and wear on the tool face and flank, while burst or transient signals result from either tool fracture or chip breakage. Therefore, the sources listed above generate both continuous AE signals and transient AE signals.

The relationship between the AE signal and tool wear is not simple. Observations of a purely progressive tool wear in turning operations indicate that in most experimental results the mean level of the averaged AE signal increases at first with an increase of flank wear, and then stays at an approximately constant level even with further increase of flank wear while the fluctuation of the mean level across the constant level becomes rather high [6]. Those observations support the conclusion that the relationship between the AE signal and the respective tool wear condition is nonlinear, so a general mathematical relation cannot be used to map this relation, [6, 9]. Effective monitoring of different tool wear conditions, while avoiding the effect of interfering factors, such as, cutting parameters is a complex task. Advanced signal processing and pattern recognition methods have been proposed as solutions to this problem, see, e.g. [6, 7], for review.

Among other indirect methods, AE is one of the most effective for sensing tool wear. The major advantage of using AE to monitor tool conditions is that the frequency range of the AE signal is much higher than that of the machine and environmental vibrations. Due to that the cutting operation essentially does not disturb the AE signals.

2.2.3 AE monitoring of the FSW process

FSW is a relatively new joining technique with continuously growing application field in joining aluminum alloys, steel, titanium and copper. However, judging from our search in the available data bases very little attention has been devoted to monitoring of the FSW process.

We are aware of only one published report concerned with using the AE technique to monitoring the FSW process. Chen at al. [2] describes an experiment involving an FSW system with a high-speed rotating tool used to weld butted plates. An on-line emission monitoring system with two AE sensors arranged symmetrically on both sides of the butting joint was used to acquire the AE signals. The sensors were moved along the welding direction with the FSW tool and the AE signals were recorded during the probe penetration, welding, and probe pullout. Since the signals acquired during the FSW experiment are non-stationary in time their time-frequency properties were analyzed using discrete wavelet transform. A wavelet transform of AE signals provided plots of frequency spectra vs time. The plots in the form of two-dimensional contour plots, showing energy as a function of frequency band and time, were found to be very useful in the recognition of the AE features.

Conclusions from the work reported above as well as the general knowledge about AE applications indicate that two major issues have to be addressed to obtain a successful AE monitoring of the FSW process:

- Suppressing irrelevant AE signals generated by external and internal sources, and
- Extracting features of AE signals that are correlated to the process variables of interest.

The first task is concerned with creating a selective AE system capable of suppressing interference generated in the FSW process including, for instance, the signals generated by the FSW machine and transmitted into the workpiece. Activity of the FSW tool generates noise that depends on its condition but also on the condition of the welded material as well as on a number of process parameters. Extracting the information on the tool condition from this mixture is a nontrivial task that requires advanced methods of signal processing.

The AE setup proposed for copper canisters

Generally, two different sources of AE signals are of interest in the FSW process of SKB's canisters: the high intensity *continuous emission* generated by the FSW tool, and the low intensity *burst emissions* due to the local releases of residual thermal stress. When both emission sources are active simultaneously their separation is a considerable challenge.

Source separation in a general case can be achieved in two ways: by filtering the received AE signals, and by using AE sensors that have selective spatial characteristics. The first way, which is commonly used, consists in designing signal filters matched to specific signal features using tools, such as, time-frequency analysis and pattern recognition. The second approach is based on using an array of AE sensors that can be steered in an electronic way. Multiple AE sensors have already been used in many AE systems for source localization. The planar localization of acoustic emission signals of the burst type is performed using the triangulation of the signal arrival times (for at least 3 probes) in order to determine the source position. Phased arrays for AE source localization were originally proposed already in 1972 by Hoff [10] and practical realization of this idea was presented by Höller at al. [11]. Lu at al. [12] explored the use of smart receiver array for locating AE sources.

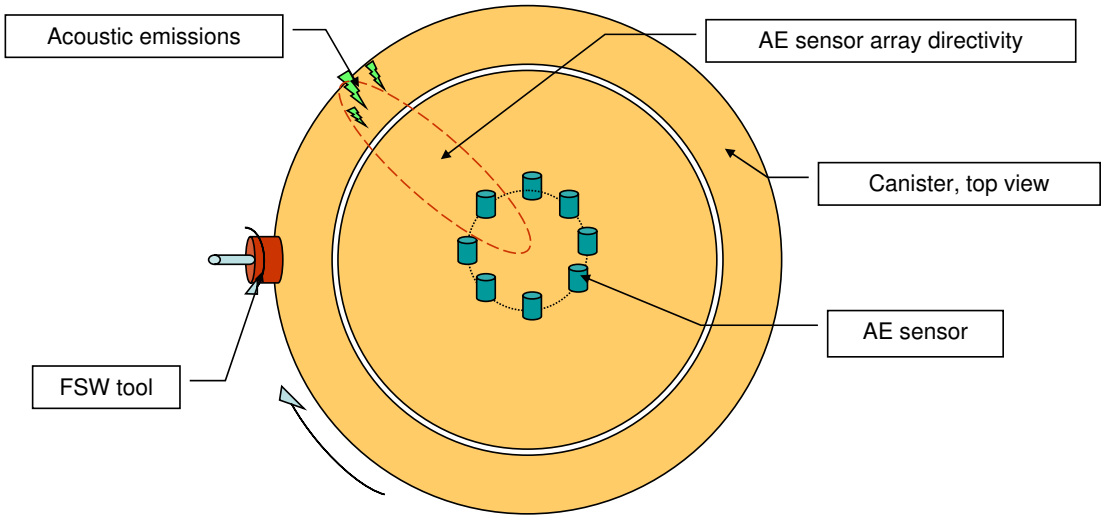


Figure 2.3: The proposed configuration of the AE monitoring system for the FSW of copper canisters.

We are proposing the test setup shown in Fig. 2.3 that takes advantage of the canister lid symmetry. A circular array of AE sensors is located in the middle of canister lid. The signals received by the individual transducers are to be fed to a beamformer capable of creating a desired beam pattern of the array in reception. The beamformer implements an algorithm consisting in coherent summation of the signals received by the individual sensors to obtain an increased sensitivity to the AE events in a pre-selected angle sector of the canister's circumference.

An off-line implementation of the beamformer using recorded signals will be tested in the first step. In the second step a hardware version of the beamformer could be implemented enabling monitoring the FSW process in real time. The primary task of the beamformer is to create a spatial-temporal window for the AE signals received by the individual AE sensors. The term 'spatial' should be understood as focusing on a pre-selected angle sector of the whole canister circumference in order to suppress signals from other AE sources. The temporal window enables focusing on a desired wave type propagating with a given velocity as well as eliminating multiple

reflected signals.

Contrary to linear uniform arrays, circular arrays are used quite seldom in ultrasound applications. Optimal beamformer design for the circular arrays requires special tools that will be briefly reviewed in the subsequent section of this report.

2.3 Uniform Circular Arrays

2.3.1 Introduction

In many scenarios for array systems, such as, wireless communications, radar and sonar, all-azimuth angle coverage is desired. One method to achieve the 360° coverage is to use uniform circular arrays (UCAs). Circular arrays have been studied extensively because of their simple structure and the unique features they offer, among those the most important is that beam shapes in the plane of the array (the azimuth plane) are essentially independent of the steering angle, [13, 14].

The desired beam shape can be achieved if the outputs of all array elements are fed to a structure, which will be referred to as beamformer in the sequel. Beamformers can take the form of analog electronic circuits (mostly in high frequency applications) or parallel algorithms for processing digitized signals from the elements. If the excitation of each array element is first theoretically determined for a desired beam shape, the conventional beamformer circuit then consists of phase shifters[1] (or delays) and attenuators (weight coefficients) inserted in all feed lines, and a combiner for all the line outputs [13].

Computationally, the phase shift and attenuation of a signal can be represented by a complex multiplication. If the signal can be sampled at a sufficiently low rate, as is the case with ultrasound signals, digital multiplications and additions can be performed in real time. In the general case, a beamformer for a circular array capable of generating a desired beam shape would have a complex structure with a considerable number of complex multiplications. The beamformer structure could be simplified considerably thanks to Butler and Lowe [15], who invented a matrix network (*Butler matrix*) of interconnected feed lines in which phase shifters serve multiple beams simultaneously. Computations according to the Butler matrix for a linear array can be completed in a much shorter time than those according to the conventional circuit.

Furthermore, it was realized that the Butler matrix was closely related to the Discrete Fourier Transform (DFT) and it could also be used to the beamforming of circular arrays. It can be shown that using the Butler matrix, any arbitrary excitation of a UCA aperture can be expressed by a spatial Fourier series. Each term in this representation constitutes an *excitation function*, referred to as a *phase mode*, whose phase changes by uniform steps from element to element as in the beam excitation of a linear array.

The Butler matrix, therefore, offers advantages in the digital beam forming process for UCA similar to those that DFT offers in the signal processing; it can serve sequentially as a spectral analyzer of broadband signals, as a mode former, and as a beamformer.

Below, we will present a brief review of the theory needed for understanding how the phase modes of a UCA can be excited using a Butler beamforming matrix.

[1] Phase shifters are equivalent to delays for narrowband signals

2.3.2 Phase Mode excitation for UCA

The UCA consists of a number of elements (transducers) uniformly distributed over the circumference of a circle of radius r in the xy plane as shown in Fig. 2.4. The individual array elements are assumed to be identical and omnidirectional. We assume that the elements are small comparing to the wavelength and their diffraction effects can be neglected.

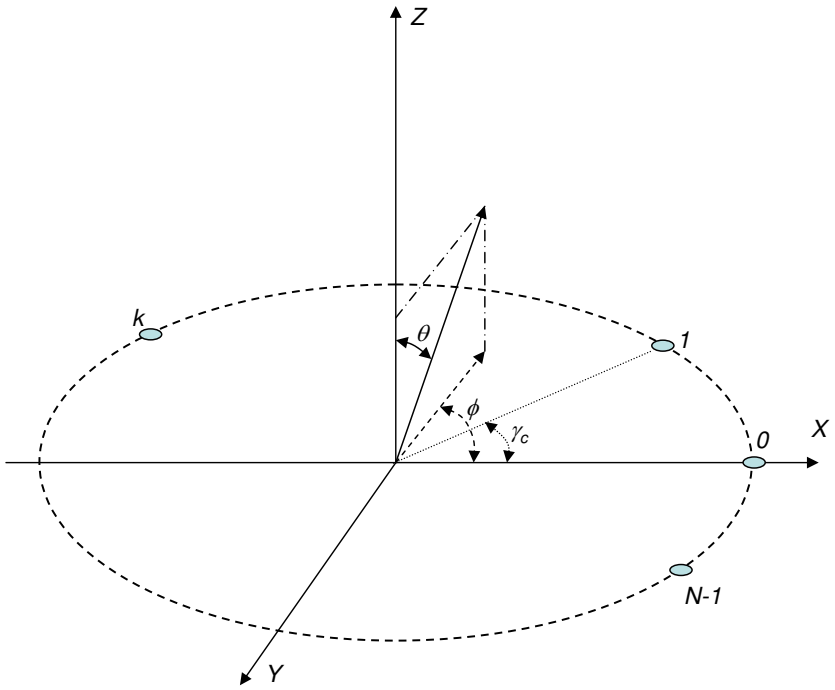


Figure 2.4: Geometry of the uniform circular array consisting of N elements.

We will use vector notation in a spherical coordinate system to represent the UCA, the beamformer, and the arrival directions of the incoming plane waves. The origin of the coordinate system is located at the center of the array and the z axis points up from the xy plane as in Fig. 2.4. Source elevation angles $\theta \in [0, \pi/2]$ are measured down from the z axis, and azimuth angles $\phi \in [0, 2\pi]$ are measured counterclockwise from the x axis.

Continuous circular aperture

Let us start from introducing the phase modes for a continuous circular aperture. Continuous circular aperture (CCA) can be thought of as a line circular source of ultrasound located at the xy plane with its center in the origin of coordinate system. The line source is excited by an *excitation function* with angular argument γ. Any excitation function $w(\gamma)$ of a CCA is periodic with period $\gamma_p = 2\pi$ and can therefore be represented using a Fourier series. The arbitrary excitation function $w(\gamma)$ thus has the representation [14]

$$w(\gamma) = \sum_{m=-\infty}^{\infty} c_m e^{jm\gamma}, \tag{2.1}$$

where c_m is the corresponding Fourier series coefficient defined as

$$c_m = \frac{1}{2\pi} \int_0^{2\pi} w(\gamma) e^{-jm\gamma} d\gamma. \tag{2.2}$$

The mth *phase mode* $w_m(\gamma) = e^{jm\gamma}$ is just a spatial harmonic of the array excitation [14].

The CCA will radiate ultrasonic waves in the xy plane as well as in the whole space. The farfield pattern in the xy plane has the same azimuthal variation $e^{jm\phi}$ as the excitation function itself.

The normalized farfield pattern for any elevation angle $\theta \neq 0$ resulting from exciting the aperture with the mth phase mode is

$$f_m^c(\mathbf{\Theta}) = \frac{1}{2\pi}\int_0^{2\pi} w_m(\gamma)e^{j\zeta\cos(\phi-\gamma)}d\zeta, \tag{2.3}$$

where the superscript c denotes the continuous aperture, $\zeta = k_0 r\sin\theta, k_0 = 2\pi/\lambda$ includes the elevational information, and the vector $\mathbf{\Theta} = (\zeta, \phi)$ represents the arrival directions of plane waves. Substituting for $w_m(\gamma)$ the elevation farfield pattern of the circular aperture can be expressed as

$$f_m^c(\mathbf{\Theta}) = j^m J_m(\zeta)e^{jm\phi}, \tag{2.4}$$

where $J_m(\zeta)$ is the Bessel function of the first kind of order m. Hence, the amplitude and elevation dependence of the farfield pattern is defined through the Bessel function $J_m(\zeta)$. Since the mode amplitude $J_m(\zeta)$ in (2.4) is small when the Bessel function order m exceeds its argument ζ, only a limited number of modes can be excited by a given circular aperture for a given wavelength. Let M denote the highest order mode that can be excited by the aperture at a reasonable strength. A rule of thumb for determining M is [14]

$$M \approx k_0 r. \tag{2.5}$$

Thus the beamformer for mode $|m| \geq M$ severely attenuates sources from all directions.

Uniform circular array

Let us consider the UCA with N elements located in the xy plane so that the first element is on x-axis as shown in Fig. 2.4. Element n of the UCA is displaced by an angle $\gamma_n = (n-1)\gamma_c = (n-1)2\pi/N$ from the x axis. The position vector at this location is $\vec{p}_n = (r\cos(n-1)\gamma_c, r\sin(n-1)\gamma_c, 0)$. Consider a narrowband plane wave with wavenumber $k_0 = 2\pi/\lambda$ propagating in the direction $-\hat{r}$, with elevation and azimuth θ and ϕ, respectively. The unit vector $\hat{r}$ has Cartesian coordinates $\hat{r} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$; the minus sign indicates the propagation direction towards the origin. The phase difference between the complex envelopes of the signals received at the origin and at element n at a given time is

$$\psi_n = e^{jk_0\hat{r}\vec{p}_n} = e^{jk_0 r\sin\theta\cos(\phi-(n-1)\gamma_c)} = e^{j\zeta\cos(\phi-(n-1)\gamma_c)},$$

where, similarly as for the CCA, $\zeta = k_0 r\sin\theta$.

We will define the *array manifold vector* $\mathbf{a}(\mathbf{\Theta})$, which incorporates all of the spatial characteristics of the UCA as

$$\mathbf{a}(\mathbf{\Theta}) = \mathbf{a}(\zeta,\phi) = \begin{bmatrix} e^{j\zeta\cos(\phi)} \\ e^{j\zeta\cos(\phi-\gamma_c)} \\ \vdots \\ e^{j\zeta\cos(\phi-(N-1)\gamma_c)} \end{bmatrix} \tag{2.6}$$

where the elevation dependence is through the parameter ζ.

Let us consider phase mode excitation of the N element UCA defined above. The normalized beamforming weight vector that excites the array with phase mode m, $|m| \leq M$ is [16]

$$\mathbf{w}_m^H = \frac{1}{N}[e^{jm0}, e^{jm\gamma_c}, \ldots, e^{jm(N-1)\gamma_c}] = \frac{1}{N}[1, e^{j2\pi m/N}, \ldots, e^{j2\pi m(N-1)/N}], \tag{2.7}$$

where $\mathbf{w}_m^H$ is a conjugate transpose of the vector $\mathbf{w}_m$.

The resulting array pattern $f_m^s(\theta)$ of a sampled circular aperture is

$$f_m^s(\mathbf{\Theta}) = \mathbf{w}_m^H \mathbf{a}(\mathbf{\Theta}) = \frac{1}{N} \sum_{n=0}^{N-1} e^{jm(n-1)\gamma_c} e^{j\zeta \cos(\phi-(n-1)\gamma_c)} \tag{2.8}$$

For mode orders $|m| < N$, the array pattern can be expressed using Bessel functions [14, 16]

$$f_m^s(\mathbf{\Theta}) = j^m J_m(\zeta) e^{jm\phi} + \sum_{q=1}^{\infty} \left(j^g J_g(\zeta) e^{-jg\phi} + j^h J_h(\zeta) e^{-jh\phi} \right) \tag{2.9}$$

where $g = Nq - 1$ and $h = Nq + m$.

Note that the first term in this equation, the principal term, is identical to the far-field pattern of (2.4) corresponding to the continuous aperture case. The remaining terms are called *residual terms*; they arise due to sampling of the continuous aperture. The principal term in (2.9) will be the dominant one if the condition $N > 2|m|$ is satisfied. The highest mode excited has order M, and

$$N > 2M \approx k_0 r \tag{2.10}$$

array elements is needed to excite all M modes [14]. This condition is identical to the Nyquist sampling criterion as M defines the maximum spatial frequency component in the array excitation. With M as in (2.10), it is clear that the circumferential spacing between adjacent array elements should be less than 0.5λ, which means that with this interelement spacing we can avoid grating lobes with UCA's. The detailed discussion concerning the contribution of the residual terms to the pattern defined by (2.9) can be found, e.g. in [14, 16].

When the residual terms in (2.9) can be ignored the array pattern for m mode can be expressed in a more compact form using the Bessel functions property $J_{-m}(\zeta) = (-1)^m J_m(\zeta)$

$$f_m^c(\mathbf{\Theta}) = j^{|m|} J_{|m|}(\zeta) e^{jm\phi}; |m| \leq M. \tag{2.11}$$

Beamforming matrices

The first step in the design of a beamformer for the UCA is to find the beamformer matrix B_{PM}^H that is based on the principle of phase mode excitation. The phase mode excitation (PME) beamformer transforms the UCA's manifold $\mathbf{a}(\mathbf{\Theta})$ to the beamspace manifold $\mathbf{a}_e(\mathbf{\Theta})$,

$$\mathbf{B}_{PM}^H \mathbf{a}(\mathbf{\Theta}) = \mathbf{a}_e(\mathbf{\Theta}). \tag{2.12}$$

The azimuthal variation of the manifold $\mathbf{a}(\mathbf{\Theta})$ produced by the beamformer (2.12) is similar to that of a uniform linear array. This is a great advantage since after this transformation the classical beamforming techniques can be used for steering the UCA resulting in compact beamformer matrices with low number of multiplications. The elevation beampattern takes the form of a symmetric taper-like main lobe and a number of side lobes.

The beamformer matrix is defined as follows [14, 16]:

$$\mathbf{B}_{PM}^H = \mathbf{C}_V \mathbf{V}^H, \tag{2.13}$$

where $\mathbf{C}_V = \text{diag}\,[j^{-M}, \ldots, j^{-1}, j^0, j^1, \ldots, j^M]$ and $\mathbf{V} = \sqrt{N}[\mathbf{w}_{-M} \vdots \ldots \vdots \mathbf{w}_0 \vdots \ldots \vdots \mathbf{w}_M]$.

The matrix $\mathbf{V}$ includes vectors $\mathbf{w}_m^H$ defined by (2.7). Each vector $\mathbf{w}_m^H$ that excites the UCA with the phase mode m generates a pattern in the form of (2.11). The term $j^{-|m|}$ in (2.11) is

canceled by the corresponding term $j^{|m|}$ in the matrix $\mathbf{C}_V$. The beamspace manifold generated by $\mathbf{B}_{PM}^H$ is thus

$$\mathbf{a}_e(\mathbf{\Theta}) = \mathbf{B}_{PM}^H \mathbf{a}(\mathbf{\Theta}) = \mathbf{C}_V \mathbf{V}^H \mathbf{a}(\mathbf{\Theta}) = \sqrt{N}\mathbf{J}_\zeta \mathbf{v}(\phi), \tag{2.14}$$

The azimuthal beam pattern of $\mathbf{a}_e(\mathbf{\Theta})$ is given by the vector

$$\mathbf{v}(\phi) = [e^{-jM\phi}, \ldots, e^{-j\phi}, e^{-j0}, e^{j\phi}, \ldots, e^{jM\phi}] \tag{2.15}$$

that defines the manifold of the uniform linear array.

The elevation dependence takes the form of a symmetric amplitude taper defined through the matrix of Bessel functions

$$\mathbf{J}_\zeta = \mathrm{diag}[J_M(\zeta), \ldots, J_1(\zeta), J_0(\zeta), J_1(\zeta), \ldots, J_M(\zeta)]. \tag{2.16}$$

The subscript e on $\mathbf{a}_e(\mathbf{\Theta})$ "even", and is intended to reflect the fact that the diagonal elements of $\mathbf{J}_\zeta$, are even about the center element.

When the beam space manifold $\mathbf{a}_e(\mathbf{\Theta})$ has been found we can operate on it to achieve a desired beam pattern using the weight vector

$$\mathbf{w}_{PM}^H = [w_{-m}^* \ldots w_0^* \ldots w_m^*]^T.$$

The output $y(k)$ of our beamformer, given the input vector $\mathbf{x}(k)$ will be

$$y(k) = \mathbf{w}_{PM}^H \mathbf{B}_{PM}^H \mathbf{x}(k). \tag{2.17}$$

The structure of the PME beamformer (2.17) is illustrated in Fig. 2.5. The signals from the UCA elements are first transformed using the matrix $\mathbf{B}_{PM}^H$, and then a desired beam pattern is created using the weight vector $\mathbf{w}_{PM}^H$ resulting in a scalar output $y(k)$.

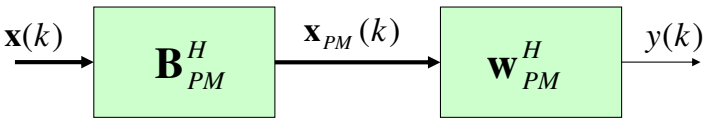
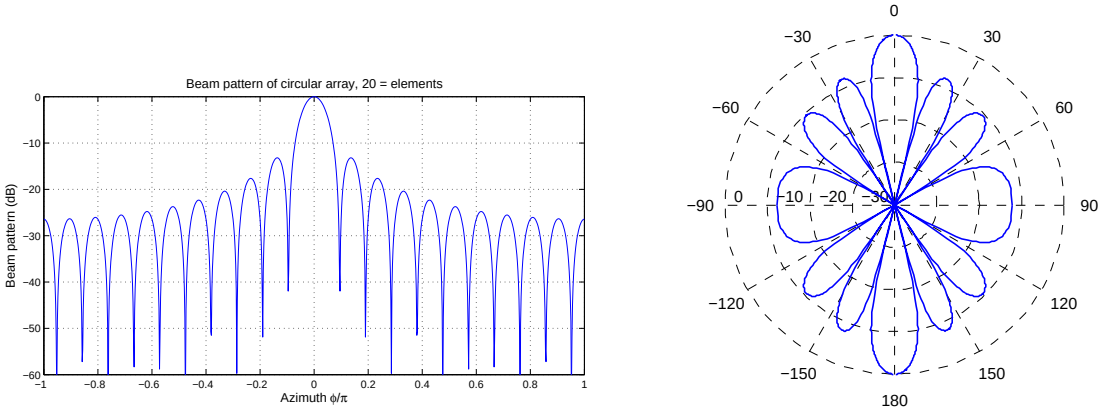


Figure 2.5: Phase mode excitation beamformer.

Example of the PME beamformer

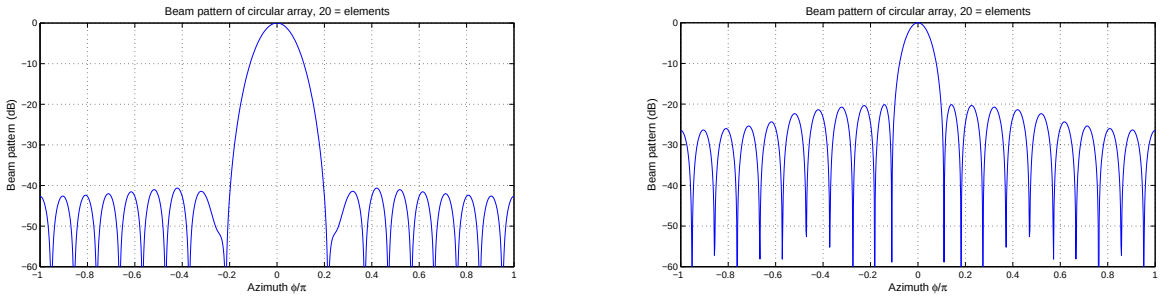
For illustration of the theory presented above we will present simulations of the theoretical beam patterns that can be obtained for the UCA with 20 elements located at the radius $R = 10\lambda/2\pi$, which results in the Nyquist spacing between the transducer equal to $\lambda/2$. For copper with sound velocity of longitudinal waves $c_l = 4660$ m/s and frequency 500 kHz the corresponding array diameter would be approx. 29.7 mm.

Both azimuthal and elevational beam patterns were simulated for different weighting vectors $\mathbf{w}_{PM}^H$. The beam patterns obtained for the uniform weights are shown in Fig. 2.6. The azimuthal pattern takes the form characteristic for uniform linear arrays with the side lobe level approx. -14dB. The elevational pattern is symmetrical with respect to the xy plane with the main lobes pointing out in both directions of the z axis.



(a) Azimuth (b) Elevation

Figure 2.6: Azimuth and elevation beam patterns of the UCA obtained with the PME beamformer with uniform weighting.

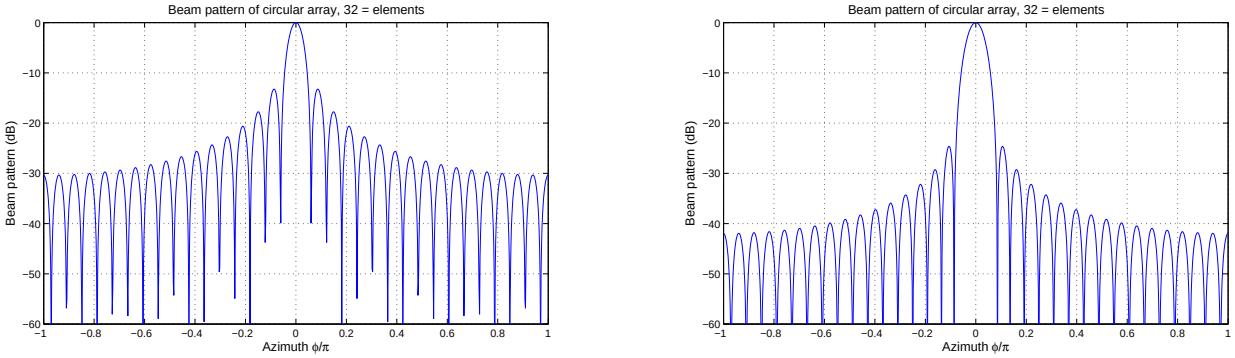


(a) Hamming window (b) Kaiser window

Figure 2.7: Azimuth beam patterns of the 20 element UCA obtained with the PME beamformer with Hamming and Kaiser window weighting, respectively.

The beam patterns obtained for weight vectors in the form of Hamming and Kaiser windows are shown in Fig. 2.7. Generally, the azimuthal patterns obtained using windowing are characterized by lower level of the side lobes obtained at the price of increased width of the main lobe.

Increased array diameter requires more elements to meet the Nyquist condition. Increasing number of elements to 32 would make possible increasing array diameter to 47.5 mm. The corresponding beam patterns are shown in Fig. 2.8.



(a) Uniform window

(b) Kaiser window

Figure 2.8: Azimuth beam patterns of the 32 element UCA obtained with the PME beamformer with uniform and Kaiser window weighting, respectively.

Note that the beam patterns presented above are correspond to a single frequency continuous wave with frequency $f = c/\lambda$.

2.4 Summary and future work

Judging from the literature, reviewed in the section 2.2, acoustic emission should be a suitable tool for monitoring the FSW process used for sealing the SKB's copper canisters. AE sensors are known to generate signals containing reach information about metal cutting processes. However, the information "richness" is an issue that has to be addressed to find suitable methods for extracting the relevant information about the tool condition.

We propose using signal processing in space and time for this purpose. Spatial signal processing consists in employing a sensor array with regular circular shape for receiving the AE signals. We have shown that such an array has nice properties, i.e., it can be steered omnidirectionally using beamformers known from the literature.

The example presented in section 2.3.2 shows that such an array can be considered as feasible; it should consist of approx. 20 elements located on a circle with diameter 30 mm if the AE signal bandwidth is limited to 0.5 MHz. Increasing the number of elements to 32 results in an array with the diameter approx. 48 mm. The array elements should be small because of two quite different reasons, first, since we are expecting waves propagating on material surface, and second, there is not much place if they are to be spaced at $\lambda/2 \simeq 4.6$ mm.

To test our concept in the laboratory we need to manufacture the UCA with suitable preamplifiers. To optimize signal to noise ratio the preamplifiers are to be integrated with the array elements. When the array is ready we will attach it to a circle plate made of aluminium. The first step of the experiment will consist in receiving artificially generated AE signals to measure

the the spatial characteristics of the UCA. The signals will be received using our 32 channel phased array system. Later we are planning to record signals generated by a drill machine with sharp and blunt drills.

Bibliography

[1] R.K. Miller and P. McIntire, editors. *NDT Handbook, Acoustic Emission Testing*, volume 5. American Society for Nondestructive Testing, 1987.

[2] C. Chen, R. Kovacevic, and D. Jandgric. Wavelet transform analysis of acoustic emission in monitoring friction stir welding of 6061 aluminum. *Int. J. of Machine Tools & Manufacture*, 43:1383–90, 2003.

[3] K. Ono. Current understanding of mechanisms of acoustic emission. *J. Strain Analysis*, 40(1), 2005.

[4] W. Swindlehurst. Acoustic emission – introduction. *Non-destructive Testing*, pages 152–58, 1972.

[5] L. M. Rogers. Structural and engineering monitoring by acoustic emission methods - fundamentals and applications. Technical report, Lloyds Register of Shipping, 2001.

[6] X. Li. A brief review: acoustic emission method for tool wear monitoring during turning. *Int. J. of Machine Tools & Manufacture*, 42:157–165, 2002.

[7] D.A. Dornfeld. Application of acoustic emission techniques in manufacturing. *NDT&E International*, 25(6):259–69, 1992.

[8] E.N. Diei and D.A. Dornfeld. A model of tool fracture generated acoustic emission during machining. *Journ. of Engineering for Industry*, 109:227–33, 2005.

[9] R.Y. Chiou and S.Y. Liang. Analysis of acoustic emission in chatter vibration with tool wear effect in turning. *Int. J. of Machine Tools & Manufacture*, 40:927–41, 2000.

[10] M. Hoff. Acoustic emission source location using single and multiple transducer arrays. *Proc. of Ultrasonics Symposium*, 1972.

[11] P. Höller, M. Klein, E. Waschkies, and G. Deuster. New approach for location of continuously emitting acoustic emission sources by phase-controlled probe arrays. *Nuclear Engineering and Design*, 128:79–82, 1991.

[12] X. Lu, W. Sachse, and I. Grabec. Use of an automatic modeler and a small receiver array for acoustic emission (AE) source location. *Ultrasonics*, 36:539–547, 1998.

[13] H.P. Raabe. Fast beamforming with circular receiving arrays. Technical report, IBM J. Res. Develop., July 1976.

[14] H.L. van Trees. *Optimum Array Processing. Part IV of Detection Estimation an Modulation Theory.* Wiley-Interscience, 2002.

[15] J. Butler and R. Lowe. Beamforming matrix simplifies design of electronically scanned antennas. *Electronic Design*, 9(12):170, April 1961.

[16] C.P. Mathews and M.D. Zoltowski. Eigenstructure techniques for $2-d$ angle estimation with uniform circular arrays. *IEEE Trans. on Sign. Proc.*, 42(9), Sept. 1994.

Chapter 3

Computational aspects of synthetic aperture imaging

by Erik Wennerström, Tadeusz Stepinski and Tomas Olofsson

3.1 Introduction

In our previous work the ESAFT image reconstruction method was developed. The ESAFT is a model based method capable of compensating for diffraction effects introduced by the imaging system using a linear model of that system. This approach has proved to yield good results in terms of spatial resolution, but at a high cost in computational complexity and workload. In practical situations, the large computational burden of the ESAFT method can become a bottle-neck in material evaluation.

In this chapter we evaluate three different approaches to perform imaging with less computational cost than of the ESAFT method. In section 3.2 a sparse version of ESAFT is presented. It utilizes the same linear model of the imaging system as the ESAFT, but solves the reconstruction problem only for a small set of the most probable scatterers in the image. In section 3.3 the ω-k SAFT algorithm is presented. It can be applied to less generic imaging problems than the ESAFT method, as it relies on the far-field approximation. It is however computationally very efficient as the bulk of the computations are done utilizing 2D Fourier transforms. Section 3.4 offers an analysis of one of the most computationally intense steps in the ESAFT and sparse 2D deconvolution methods. A scheme to reorder calculations that considerably reduces the computational burden is presented.

3.2 Sparse deconvolution of B-scan images

3.2.1 Introduction

In this chapter, a new computationally efficient sparse deconvolution algorithm for the use on B-scan images from objects with relatively few scattering targets is presented. It is based on a similar linear image formation model that has been used earlier in the ESAFT algorithm. The linear minimum mean squared error (MMSE) 2-d deconvolution results obtained with ESAFT have shown high resolution, but at the cost of increased computation time.

The proposed algorithm utilizes the sparsity of the image, reducing the degrees of freedom in the reconstruction problem, to reduce the computation time and to improve the resolution. The dominating task in the algorithm consists in detecting the set of active scattering targets, which is done by iterating between one up-dating pass that detects new points to include in the set, and a down-dating pass that removes redundant points. In the up-date, a spatio-temporal matched filter is used to isolate potential candidates. A subset of those are chosen using a detection criterion. The amplitudes of the detected scatterers are found by the MMSE deconvolution. The algorithm properties are illustrated using synthetic and real B-scans and the involved computation times are analyzed.

Resolution enhancement of ultrasonic pulse-echo B-scan images can be achieved using SAFT or using estimation based approaches, such as 2-d deconvolution techniques. In the estimation based approaches, the entities of interest are the scattering strengths from points in the image plane. Although we may argue that SAFT techniques also aim at finding this map of scattering strengths, these methods do not use any statistically motivated method for estimation; they should rather be interpreted as mimicking the behavior of an acoustical lens using the delay-and-sum (DAS) technique. As such, they inherit the intrinsic limitations of DAS; the lateral resolution is diffraction limited and grating lobes appear when using too coarse a spatial sampling

[1].

Methods using the estimation based approach can, for images with low noise levels, overcome the above mentioned limitations and far outperform SAFT. For mechanically scanned B-scans, examples of this class of methods can be found in [2] where linear MMSE was used to estimate the scatter map. A similar method based on singular value decomposition (SVD) regularization is found in [3]. A serious disadvantage of the methods is that they require computation time that is typically much longer than that for SAFT processing, often limiting their use only to applications without real-time constraints.

Linear methods as those in [2] and [3] are best adapted to objects containing diffuse scatterers, with a Gaussian distribution modeling the amplitudes of these. In many application, particularly in nondestructive testing (NDT), the images are better described by a collection of a few but relatively strong contribution, for instance, indicating small cracks or material inclusions. For such images, sparse models and sparse deconvolution techniques are more appropriate. The majority of work on sparse deconvolution has concerned 1-d signals. Sparse deconvolution can be described as the combined detection of the spikes and estimation of their amplitudes.

The presence of overlapping contributions from the individual scatterers makes the detection of their images (spikes in sparse deconvolution) a complicated task. Iterative search approaches have been used to find the most likely combination of the spike positions. The single most likely replacement (SMLR) algorithm [4] has been widely used for this purpose in both the Kalman and linear algebra approaches. SMLR iteratively scans every potential spike position and finds the replacement yielding the largest improvement in the chosen criterion. The replacement may consist in either including or removing a point from a set of positions found so far.

One of relatively few examples of sparse deconvolution in imaging applications can be found in star field astronomy [5] where a spatially invariant point spread function (PSF) well describes the image formation. Note, however, that the linear algebra approach mentioned above allows for treating any linear measurement model, where image generation models with spatially variant PSFs are important special cases. A recent paper by the authors [6] proposed the use of 2-d sparse deconvolution for ultrasonic B-scan images in which such spatially varying PSFs were considered. Note that the involved matrices and vectors are usually much larger in imaging applications than in 1-d deconvolution. This is particularly true for the spatially variant PSFs relevant to mechanically scanned B-scans and time- and memory efficient methods are therefore much desired.

In [6] in was shown that the sparsity of the image could be utilized to significantly reduce the time required for the estimation compared to linear MMSE solutions. An efficient search strategy was developed that can be regarded as a 2-d extension of the multiple most likely replacement (MMLR) search in [7], which was based on recursions that are not applicable for images. In [6], the search strategy was combined with efficient computations of the search criterion and a simulation study verified that significant time saving were possible. However, few algorithm details were given and no results were presented for real data. This paper gives a more thorough presentation of the algorithm and presents results for real B-scans.

The Section 3.2 is organized as follows: In Section 3.2.2, the image formation model and the search criterion are presented. In Section 3.2.3, the algorithm is presented along with some of the related computational issues. Simulation- and experimental results are presented in Section 3.2.4. Finally, in Section 3.2.5 conclusions and comments are given.

3.2.2 Theory

Image model

Consider the measurement setup depicted in Fig. 3.1 in which a transducer is scanned in the x-direction and transmits in the z-direction. At each of the L scanning positions, a pulse echo measurement is performed insonifying the region of interest (ROI). The ROI, which lies in the zx-plane, consists of a rectangular grid of $M \times N$ equally spaced potential scatterers. The spacing is Δx and Δz in x- and z directions, respectively. Furthermore, Δx is set equal to the distance between the scanning positions. Although nothing prevents from choosing a quite arbitrary grid, this particular choice results in simplified algorithm implementations.

The received A-scans are sampled at K time instants with a sampling period T_s. We have chosen $\Delta z = T_s c_p/2$, where c_p is the sound speed. This corresponds to one grid point per time sample.

Let $y(k,l)$ denote the lth A-scan, sampled at time k and let $\mathbf{y}_l = col(y(1,l), y(K,l), ..., y(K,l))$ represent this A-scan as a $K \times 1$ vector. L such column vectors are acquired and the B-scan can be represented by a column vector $\mathbf{y} = (\mathbf{y}_1^T ... \mathbf{y}_L^T)^T$, i.e., a column vector of size $KL \times 1$ consisting of the A-scans placed on top of each other.

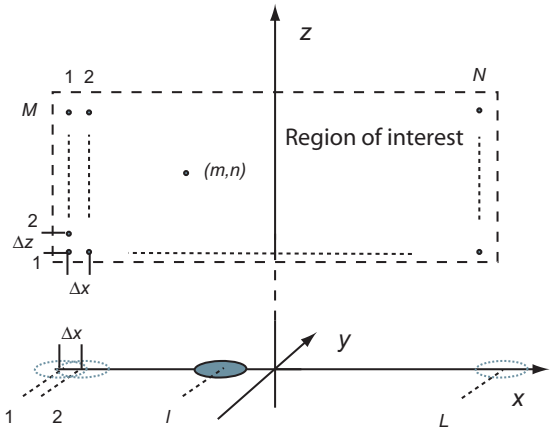


Figure 3.1: Illustration of the measurement setup in the text. The ROI is divided in an $M \times N$ grid with the center points in the grid separated with Δx and Δz in the x- and z-directions, respectively. Pulse echo measurements are performed at L transducer positions with a scan increment of Δx, resulting in a B-scan of size $K \times L$.

Let further the column vector $\mathbf{p}_{(m,n)}$ denote the vectorized B-scan that a hypothetical (noise free) response from a unit strength scatter at grid position (m,n) would result in. This template is defined by the electro-mechanical impulse response of the transducer and the double-path spatial impulse responses (SIRs) associated with the scatterer position, see [2] for details.

We consider the B-scan to be the result of backscattering from a subset, $\mathcal{I}$, of the points in the ROI. In the following we denote this subset as the set of *active scatterers*. Let n_a denote the number of points in this set.

If we neglect multiple scattering, we can model the B-scan as a noise-corrupted superposition of contributions from active scatterers,

$$\mathbf{y} = \sum_{(m,n)\in\mathcal{I}} \mathbf{p}_{(m,n)} o(m,n) + \mathbf{e} = \mathbf{P}_{\mathcal{I}} \mathbf{o}_{\mathcal{I}} + \mathbf{e}, \tag{3.1}$$

where the sum runs over only the positions listed in $\mathcal{I}$. The matrix $\mathbf{P}_{\mathcal{I}}$ consists of n_a columns $\mathbf{p}_{(m,n)}$ with $(m,n) \in \mathcal{I}$ and the $n_a \times 1$ vector $\mathbf{o}_{\mathcal{I}}$ consists of the corresponding amplitudes

$o(m,n)$. Finally, the vector $\mathbf{e}$ represents measurement noise. In practice, we let this vector model anything that cannot be explained by the linear relation in (3.1). This could include model discrepancies, such as errors in the template $\mathbf{p}_{(m,n)}$ predicting the response from a scatterer at position (m,n). This can, for instance, occur when using an inappropriate method for the spatial impulse response (SIR) calculation, errors in the measurement of the transducer impulse response, or because the true scatterer positions are not found exactly on the grid points.

Statistical assumptions

The image is modeled using a Bernoulli-Gaussian distribution. A priori, a scatterer is independently of the other scatterers assumed to belong to the active set with a probability λ. This number roughly describes how large a fraction of the grid positions that can be expected to contain an active scatterer.

We model the noise as zero mean white Gaussian, $\mathbf{e} \sim N(0, \sigma_e^2 \mathbf{I}_{KL})$, where $\mathbf{I}_{KL}$ denotes a $KL \times KL$ identity matrix. Furthermore, we model the amplitudes of the active scatterers to be independent and identically distributed (iid) as $o_m \sim N(0, \sigma_o^2)$, yielding the joint distribution $\mathbf{o}_\mathcal{I} \sim N(0, \sigma_o^2 \mathbf{I}_{n_a})$. If necessary, these assumptions can be relaxed to include colored noise as well as correlations between elements in $\mathbf{o}_\mathcal{I}$.

Search criterion

We define sparse deconvolution as the combination of (i) joint detection, i.e., finding the set $\mathcal{I}$ that maximizes the probability $P(\mathcal{I}|\mathbf{y})$, which is equivalent to maximizing the product $p(\mathbf{y}|\mathcal{I})P(\mathcal{I})$, and then, as a final stage, (ii) estimating the amplitudes indicated by the so found $\mathcal{I}$. Step (i) is the dominating task and is solved iteratively by successively improving a criterion function that is obtained as follows:

We take as the criterion the logarithm of the above product, with constant terms removed. The logarithm of $P(\mathcal{I})$ is given by

$$\ln P(\mathcal{I}) = \ln(\lambda^{n_a}(1-\lambda)^{NM-n_a}) = n_a \ln \frac{\lambda}{1-\lambda} + const. \tag{3.2}$$

To find $\ln p(\mathbf{y}|\mathcal{I})$, the marginalization integral is solved yielding

$$p(\mathbf{y}|\mathcal{I}) = \int p(\mathbf{y}, \mathbf{o}_\mathcal{I}|\mathcal{I}) d\mathbf{o}_\mathcal{I} = \frac{exp(-\frac{1}{2}\mathbf{y}^T \mathbf{D}_\mathcal{I}^{-1}\mathbf{y})}{(2\pi)^{NM/2}|\mathbf{D}_\mathcal{I}|^{1/2}}, \tag{3.3}$$

where $\mathbf{D}_\mathcal{I}$ is given by

$$\mathbf{D}_\mathcal{I} = \sigma_e^2 \mathbf{I}_{KL} + \sigma_o^2 \mathbf{P}_\mathcal{I} \mathbf{P}_\mathcal{I}^T. \tag{3.4}$$

This matrix is of size $KL \times KL$, which makes the quadratic form and determinant in (3.3) computationally awkward. Both these factors can however be simplified to more attractive forms which results in the log marginal likelihood

$$\begin{aligned} \ln p(\mathbf{y}|\mathcal{I}) =& n_a \ln \beta - \frac{1}{2} \ln \left| \mathbf{B}_\mathcal{I} + \beta^2 \mathbf{I}_{n_a} \right| - \frac{1}{\sigma_e^2} ||\mathbf{y}||^2 \\ &+ \frac{1}{2\sigma_e^2} \mathbf{c}_\mathcal{I}^T (\mathbf{B}_\mathcal{I} + \beta^2 \mathbf{I}_{n_a})^{-1} \mathbf{c}_\mathcal{I}, \end{aligned} \tag{3.5}$$

where $\beta = \frac{\sigma_e}{\sigma_o}$, where the $n_a \times 1$ vector $\mathbf{c}_I$ is defined as

$$\mathbf{c}_I = \mathbf{P}_\mathcal{I}^T \mathbf{y}, \tag{3.6}$$

and with the symmetric $n_a \times n_a$ matrix $\mathbf{B}_\mathcal{I}$ defined as

$$\mathbf{B}_\mathcal{I} = \mathbf{P}_\mathcal{I}^T \mathbf{P}_\mathcal{I}. \tag{3.7}$$

Note that the elements in $\mathbf{B}_\mathcal{I}$ are scalar products between the different $\mathbf{p}_{(m,n)}$-vectors in the active set. These can be extracted from a matrix (look-up-table) $\mathbf{B} = \mathbf{P}^T\mathbf{P}$, where $\mathbf{P}$ is the $KL \times MN$ matrix consisting of the $\mathbf{p}$-vectors for all points in the ROI. Note also that $\mathbf{B}$ does not depend on the received data and it can be computed and stored off-line.

Similarly, the vector $\mathbf{c}_\mathcal{I}$ for a certain $\mathcal{I}$ can be extracted from a $MN \times 1$ vector $\mathbf{c} = \mathbf{P}^T\mathbf{y}$, that we can compute before we begin the search for the optimal $\mathcal{I}$. The vector $\mathbf{c}$ is the output from a spatio-temporal matched filter and large entries in $\mathbf{c}$ indicates the likely presence of an active scatterer. This is used in the search strategy described in Section 3.2.3.

By combining (3.2) and (3.5) and removing terms that are constant with respect to $\mathcal{I}$, we obtain the search criterion

$$\begin{aligned} J(\mathcal{I}) =& n_a\alpha - \frac{1}{2}\ln\left|\mathbf{B}_\mathcal{I} + \beta^2\mathbf{I}_{n_a}\right| \\ &+ \frac{1}{2\sigma_e^2}\mathbf{c}_\mathcal{I}^T(\mathbf{B}_\mathcal{I} + \beta^2\mathbf{I}_{n_a})^{-1}\mathbf{c}_\mathcal{I}, \end{aligned} \tag{3.8}$$

with the scalar $\alpha = \ln\frac{\lambda}{1-\lambda} + \ln\beta$.

Note also that for a given $\mathcal{I}$, the linear MMSE estimate of the amplitudes is given by

$$\hat{\mathbf{o}}_\mathcal{I} = (\mathbf{B}_\mathcal{I} + \beta^2\mathbf{I}_{n_a})^{-1}\mathbf{c}_\mathcal{I}. \tag{3.9}$$

The linear (non-sparse) MMSE solution used in [2] is obtained as the special case when $\mathcal{I}$ consists of all points in the ROI:

$$\hat{\mathbf{o}}_{MMSE} = (\mathbf{P}^T\mathbf{P} + \beta^2\mathbf{I}_{MN})^{-1}\mathbf{P}^T\mathbf{y}. \tag{3.10}$$

3.2.3 Algorithm

Initialization

The initialization consists of calculating the look-up-tables $\mathbf{B}$ and $\mathbf{c}$ and for this we first need to calculate the $\mathbf{p}$-vectors. These are defined by the transducer's electrical impulse response, which is measured, and the SIRs associated with the points in the ROI. In this work we have used the DREAM (Discrete REpresentation Array Modeling) toolbox [8, 9] for calculating the SIRs.

Note that choosing the grid point distance and scanning step to be identical results in that the response from a scatterer (m, n) at scan positions l is identical to a response from $(m+1, n)$ at $l+1$. As a result from this shift invariance, the matrix $\mathbf{P}$ is block Toeplitz containing $L \times N$ blocks, each of size $K \times M$. Only $M + N$ blocks need to be calculated to fully determine $\mathbf{P}$.

We have developed an algorithm that utilizes this block Toeplitz structure in the computation of $\mathbf{B}$. Matrix products of the smaller blocks occur at several instants in $\mathbf{B}$ and by avoiding repeated calculations of these block we reduce the computation time from being of order $(KM)^2 \times (LN)^2$ to $(KM)^2 \times max(L, N)^2$.

Search

We aim at, as efficiently as possible, finding the set of $\mathcal{I}$ that maximizes the criterion $J(\mathcal{I})$. In the search we generate a sequence of sets that yield monotonically increasing $J(\mathcal{I})$ and the

search ends when no further increase is achieved. We start at iteration $i = 0$ with the empty set $\mathcal{I}^0 = \emptyset$.

At iteration i, the search has produced the set $\mathcal{I}^i$. The next iteration then consists of the following steps:

1. Adding elements (up-dating). The elements to add are chosen as follows:
   (a) Compute the output from the matched filter applied to a residual image, $\mathbf{c}_{res} = \mathbf{P}^T\mathbf{y}_{res}$ where $\mathbf{y}_{res} = \mathbf{y} - \mathbf{P}_{\mathcal{I}^i}\hat{\mathbf{o}}_{\mathcal{I}^i}$. $\hat{\mathbf{o}}_{\mathcal{I}^i}$ is the vector of estimated amplitudes for the scatterers indicated in $\mathcal{I}^i$. Intuitively, $\mathbf{y}_{res}$ is the remaining B-scan after having removed the effects of the so far found scatterers. By applying the matched filter, we try to detect scatterers that best help to explain the residual B-scan.
   (b) Find a set of candidates, $\mathcal{C}$, by extracting peaks in the point-wise squared matched filter output $|\mathbf{c}_{res}|^2$ and including also their neighboring points in $\mathcal{C}$. These neighboring points are found by a user-defined mask centered at each peak.
   (c) Compute $J(\mathcal{I}^i \cup (m,n))$ for all points $(m,n) \in \mathcal{C}$. Remove all points from $\mathcal{C}$ that do not yield improvements compared to $J(\mathcal{I}^i)$.
   (d) Iterate until $\mathcal{C}$ is empty: (i) pick the currently best point $(m^*,n^*) = \arg\max_{(m,n)\in\mathcal{C}} J(\mathcal{I}^i \cup (m,n))$. (ii) remove from $\mathcal{C}$ all elements that are not orthogonal to $\mathbf{p}_{(m^*,n^*)}$. Let $\mathcal{U}$ denote the set of points found in this iteration.
2. Removing redundant elements (down-dating). For all $(m,n) \in (\mathcal{I}^i \cup \mathcal{U})$, calculate $J((\mathcal{I}^i \cup \mathcal{U}) \setminus (m,n))$, where '\' denotes set difference. Remove those elements that yield improvements compared to $J(\mathcal{I}^i \cup \mathcal{U})$.

Step (1-d) relies on the following: Suppose that we have several candidates that all yield increases in J if they are included in $\mathcal{I}^i$. Unfortunately, there is no guarantee that these elements will yield an increase in J if they are *simultaneously* included in $\mathcal{I}^i$. However, it can be shown that if their associated $\mathbf{p}$-vectors are all mutually orthogonal, the increase in J that is obtained by including all candidates can be written as a sum of the individual increases. This sum is guaranteed to be positive since all these terms are positive.

The $\mathbf{p}$-vectors describe contributions from different scatterers. Even for well separated points, there will be a slight overlap of these contributions in the B-scans and the corresponding vectors will therefore typically not be perfectly orthogonal. Since, for all practical purposes, it is usually sufficient that the scalar product is small, but not necessarily zero, we say that the vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ are orthogonal if $|\mathbf{p}_1^T\mathbf{p}_2| < tol$, where tol is a user-defined tolerance.

Computational aspects

For a fast search we need efficient methods to evaluate J. The majority of the evaluations occur in up-date (1-c) and down-date (2) and different computation strategies are suitable in these cases. In the up-date we base the computations on lower triangular Cholesky factors of $\mathbf{B}_{\mathcal{I}} + \beta^2\mathbf{I}_{n_a}$, satisfying

$$\mathbf{G}_{\mathcal{I}}\mathbf{G}_{\mathcal{I}}^T = \mathbf{B}_{\mathcal{I}} + \beta^2\mathbf{I}_{n_a}. \tag{3.11}$$

This factorization requires approximately $n_a^3/3$ floating point operations (flops). Having $\mathbf{G}_{\mathcal{I}}$ we can then compute J as

$$J(\mathcal{I}) = n_a\alpha - \sum_{k=1}^{n_a} \ln(\mathbf{G}_{\mathcal{I}})_{k,k} + \frac{1}{2\sigma_e^2}||\mathbf{v}_{\mathcal{I}}||^2, \tag{3.12}$$

where $(\mathbf{G}_\mathcal{I})_{k,k}$ is the kth diagonal element of $\mathbf{G}_\mathcal{I}$ and $\mathbf{v}_\mathcal{I}$ is the solution to the equation $\mathbf{G}_\mathcal{I}\mathbf{v}_\mathcal{I} = \mathbf{c}_\mathcal{I}$. For moderate to large n_a, the computations in (3.12) require approximately $2n_a + n_a^2/2$ flops. Note further that $\hat{\mathbf{o}}_\mathcal{I}$ in eq. (3.9) can be obtained by solving $\mathbf{G}_\mathcal{I}^T\hat{\mathbf{o}}_\mathcal{I} = \mathbf{v}_\mathcal{I}$. Thus, having computed $\mathbf{v}_\mathcal{I}$, approximately $n_a^2/2$ flops are required to estimate the amplitudes.

These estimated amplitudes are used in step (1-a) for evaluating $\mathbf{c}_{res} = \mathbf{P}^T\mathbf{y}_{res}$. It can be written as $\mathbf{c}_{res} = \mathbf{c} - \mathbf{P}^T\mathbf{P}_{\mathcal{I}^i}\hat{\mathbf{o}}_{\mathcal{I}^i}$ where $\mathbf{c}$ has been pre-computed and all elements in the product $\mathbf{P}^T\mathbf{P}_{\mathcal{I}^i}$ can be directly extracted from the matrix $\mathbf{B}$.

In the up-date in (1-c) we evaluate J for a number of related sets. We have already computed $\mathbf{G}_\mathcal{I}$ and $\mathbf{v}_\mathcal{I}$ in an earlier iteration and we can then find the values for these sets using a recursive up-date of the Cholesky factor that results in the following up-date equations for obtaining $J(\mathcal{I}^i \cup (m,n))$:

$$\begin{aligned} J(\mathcal{I}^i \cup (m,n)) =& J(\mathcal{I}^i) + \alpha - \ln q \\ &+ (\mathbf{p}_{(m,n)}^T\mathbf{y} - \mathbf{z}^T\mathbf{v}_\mathcal{I})^2/(2q^2\sigma_e^2), \end{aligned} \tag{3.13}$$

where $\mathbf{z}$ is the solution to $\mathbf{G}_\mathcal{I}\mathbf{z} = \mathbf{b}_{(m,n)}$, with $\mathbf{b}_{(m,n)}$ being a vector of scalar products between $\mathbf{p}_{(m,n)}$ and those $\mathbf{p}$-vectors corresponding to the elements in $\mathcal{I}^i$. All these are available in $\mathbf{B}$. The scalar product $\mathbf{p}_{(m,n)}^T\mathbf{y}$ is found in $\mathbf{c}_\mathcal{I}$ and finally, $q = ||\mathbf{p}_{(m,n)}||^2 + \beta^2 - ||\mathbf{z}||^2$, where the first term, again, is available in $\mathbf{B}$. The computations in (3.13) involve approximately $n_a + n_a^2/2$ flops.

In the down-dating we calculate $\hat{\mathbf{o}}_\mathcal{I}$ as explained earlier and explicitly compute the inverse $\mathbf{D}_\mathcal{I} = (\mathbf{B}_\mathcal{I} + \beta^2\mathbf{I}_{n_a})^{-1}$, which can be efficiently done using $\mathbf{G}_\mathcal{I}$ available from an earlier up-date step. Let d denote the element on the diagonal in $\mathbf{D}_\mathcal{I}$ corresponding to point (m,n). The following equations can then be used:

$$J(\mathcal{I} \setminus (m,n)) = J(\mathcal{I}) - \alpha - \frac{1}{2}\ln d - \frac{\hat{o}(m,n)^2}{2\sigma_e^2 d}, \tag{3.14}$$

where $\hat{o}(m,n)$ is the estimated amplitude at point (m,n).

3.2.4 Simulations and experimental results

Experimental setup

The performance of the proposed algorithm was evaluated using simulated and measured data. The measurements were performed in immersion and 0.3 mm thin steel wires were used to simulate point-like targets. A water sound speed of 1480 m/s was assumed in the experiment. The data was acquired using a circular 2.25 MHz piston transducer with 6 mm radius. The spatial sampling distance was Δx=1 mm and temporal sampling frequency was 25 MHz. The targets were located at a distance of 120 mm from the transducer and were grouped in three pairs, in which the targets were separated by a distance of 5, 3 and 1 mm, respectively. This separation was in the scanning direction x as illustrated in Fig. 3.2.

The considered ROI had the size 5 mm $\times$ 70 mm yielding $160 \times 70 = 11200$ grid points. The corresponding B-scan had the same size.

The wire targets were chosen to avoid problems with too weak signals when using small point-like reflectors. However, the use of these violates the point target assumption; the response from the wires can be approximated by an integral over the points on a line, which corresponds to a temporal smearing effect of the signals compared to smaller targets. Although we can re-define our model to include this effect, we chose in this work the approximate but simpler method of

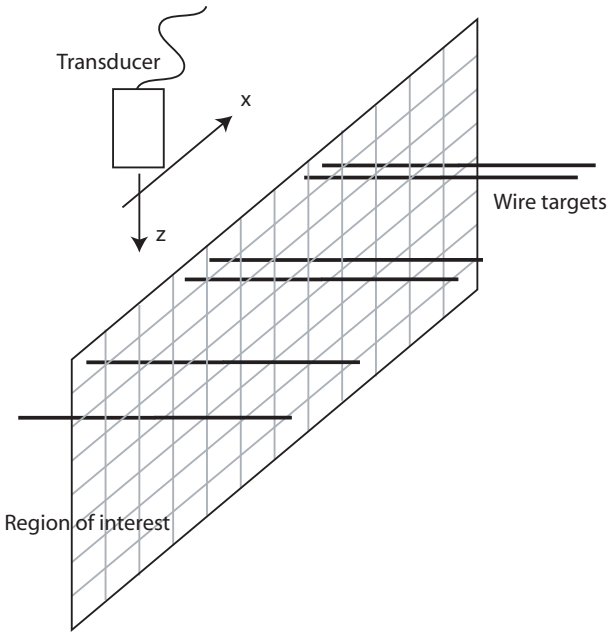


Figure 3.2: Measurement setup. Wire targets in immersion.

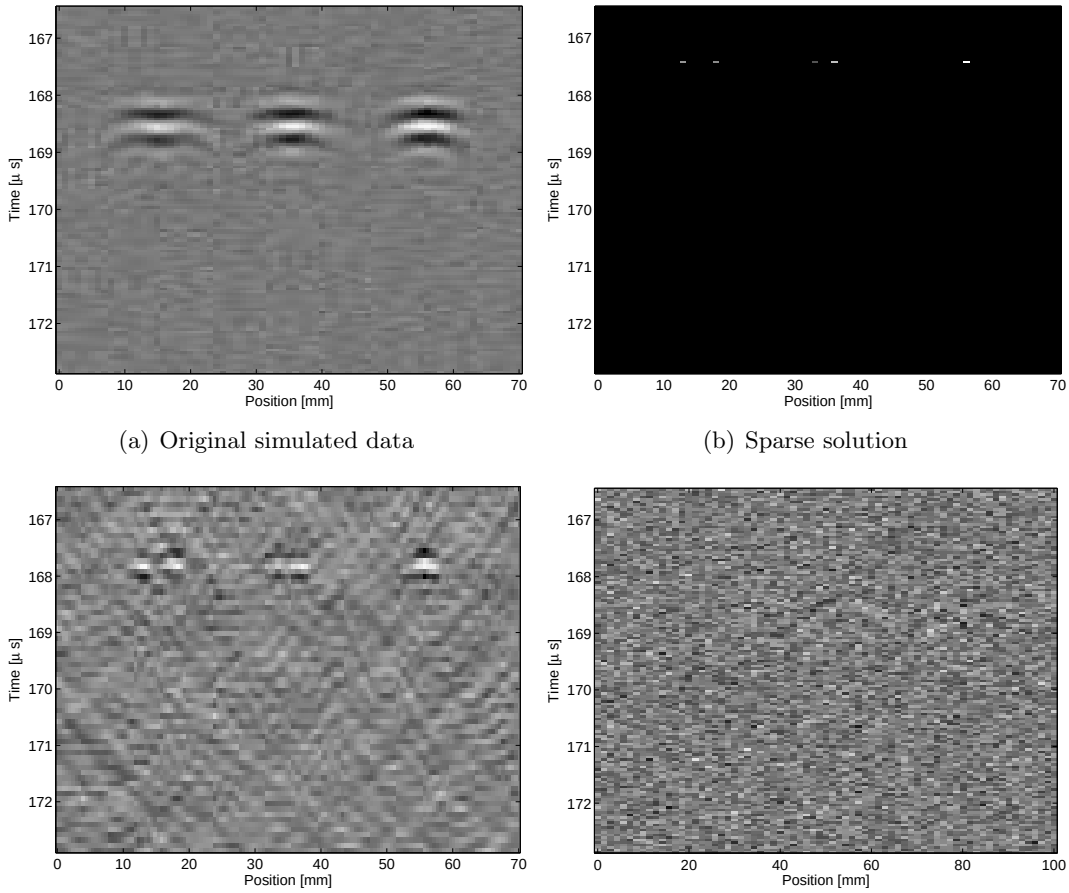
including the smearing directly in the transducer impulse response by measuring this impulse response as the pulse-echo signal obtained from a wire target in the far field and deconvolving this measured response with the SIR associated with the on-axis point at the same distance.

Simulation results

Measurements from the above setup were simulated. The SIRs used in the simulations were obtained using the DREAM toolbox [8] and the transducer impulse response was the same as in the real data experiments. The scattering amplitudes at the wire positions were set to 3×10^{-8} which yielded a peak-to-peak amplitude of 300 quantization levels.[1] White Gaussian noise of variance $\sigma_e^2 = 100$ was added to the simulated data to obtain an SNR of approximately 10 dB.

The results from the simulations are presented in figures 3.3(a) to 3.3(c). Figure 3.3(a) shows the simulated raw data. Figure 3.3(b) shows the sparse solution obtained with parameters $\sigma_e^2 = 100$, $\sigma_o^2 = 10^{-15}$, and $\lambda = 0.01$. In figure 3.3(c), a corresponding linear MMSE solution [2] obtained using eq. (3.10) is shown. To obtain comparable results in the sense that the models predict approximately the same energy in the images, the prior variance of amplitudes was set to, $\sigma_o^2\lambda = 10^{-17}$ in the linear MMSE solution. Finally, the residual for the sparse solution, $\mathbf{y}_{res} = \mathbf{y} - \mathbf{P}_I\hat{\mathbf{o}}_I$, is displayed in figure 3.3(d).

[1] The choice of scatterer amplitudes are arbitrary as long as they, in combination with the SIRs and the transducer impulse response, yield signals of realistic magnitudes. We may choose to up-scale the amplitudes and down-scale the impulse response the same amount without any effect on the detection results.



(a) Original simulated data

(b) Sparse solution

(c) Linear MMSE solution

(d) Residual for the sparse solution

Figure 3.3: Results from simulations

We see in the figures that the sparse algorithm can distinguish the point target pairs that are separated by 5 mm and 3 mm. The pair of targets that are separated with only 1 mm are however estimated as a single, but stronger scatterer. As a comparison, the linear MMSE can well resolve the pair separated with 5 mm well whereas a separation of next pair of targets is only faintly indicated. No separation of the 1 mm targets can be seen.

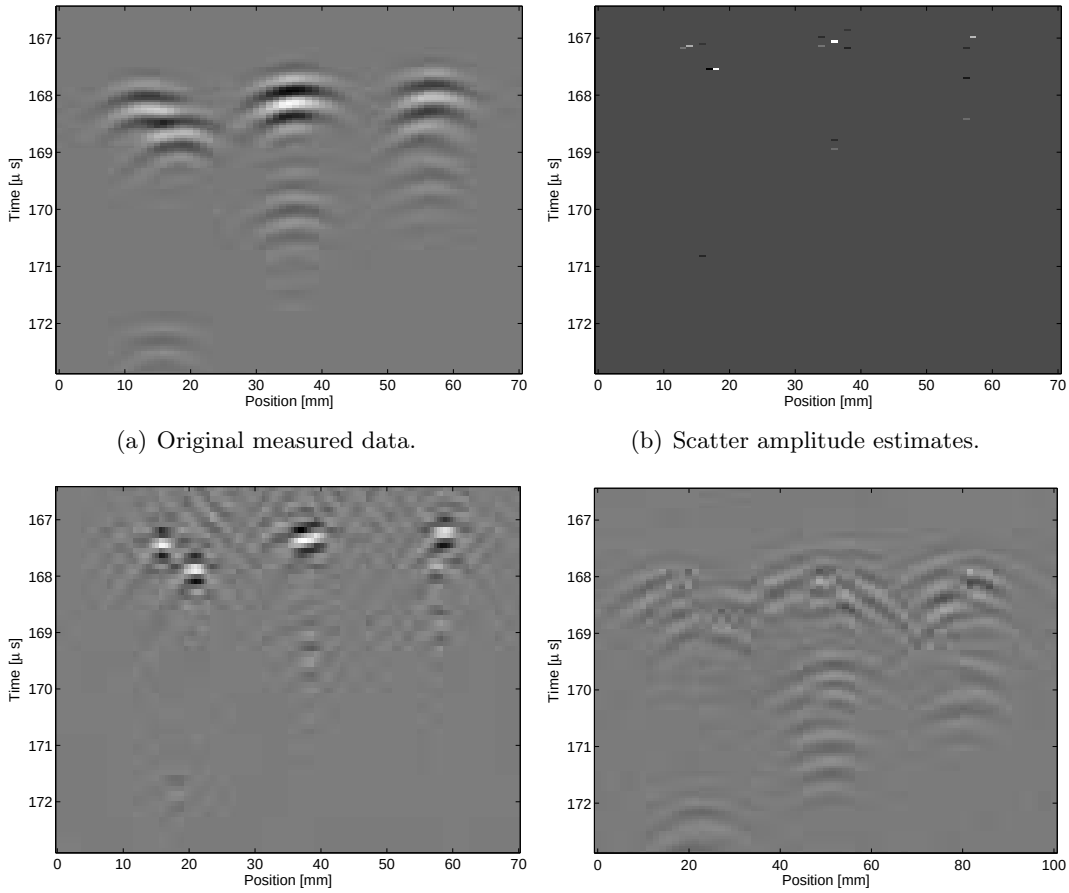
Except for the last error, the detection of the positions in the sparse solution is perfect and the amplitude estimation errors are less than 10% of the true values.

The residual of the sparse solution shows mostly white noise. Most of the energy in the original data has been accounted for which could be expected since no model errors are present; the simulated data is generated from the same model used for reconstruction.

The simulation results serve as an example of the performance that can be achieved using sparse deconvolution in an ideal situation with no model errors present. We note that excellent resistance to noise and the detection capabilities can be achieved under such circumstances.

Experimental results

The results from processing of the measured data is displayed in figure 3.4(a) to 3.4(c).



(a) Original measured data.

(b) Scatter amplitude estimates.

(c) MMSE result.

(d) Residual B-scan.

Figure 3.4: Results from experimental data

The wires can be seen as the white spots, corresponding to high positive estimated scattering amplitudes. Weaker dark indications[2] near the main echoes are most probably caused by model errors that, for instance, may stem from slightly inaccurate SIRs or targets that not residing exactly at the grid positions. As can be seen in the figures, the wires were not positioned on a perfectly horizontal line. The leftmost wire in the first pair was slightly closer to the transducer than the other.

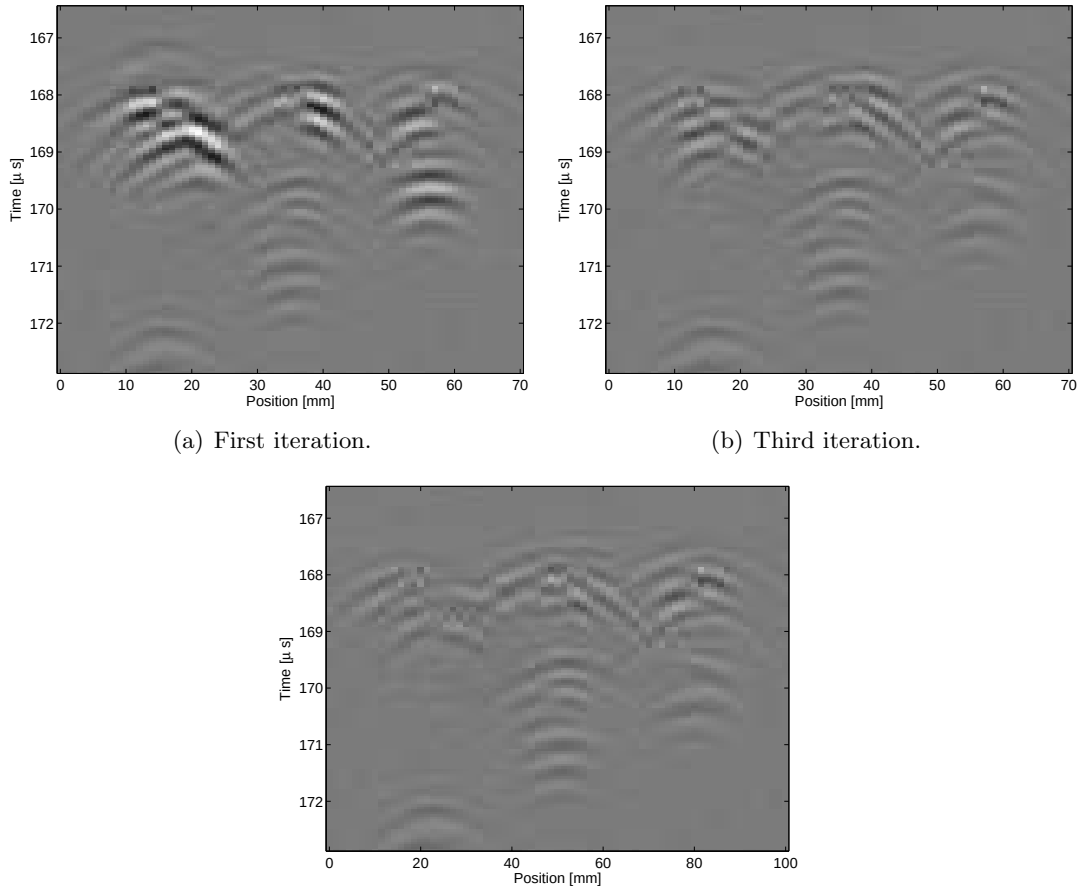
Similar to the simulations, the wires that were separated with 5 mm and 3 mm can be distinguished by the sparse algorithm, but the two wires that were 1 mm apart are estimated as a single, stronger scatterer. The corresponding linear MMSE solution can now clearly separate only the wires in the first pair.

In the measured data we can observe additional contributions that can be explained by the sound paths: transducer-wire-wire-transducer. These contribution appear approximately 3.5 μs, 1.8 μs, and 0.8 μs after for main echoes, for the wires separated by 5 mm, 3 mm, and 1 mm, respectively. This agrees fairly well with the predicted values 3.4 μs, 2 μs, and 0.7 μs predicted by a sound propagation speed of $c_p = 1480$ m/s.

To illustrate the updating process, figures 3.5(a) to 3.5(c) show the residual B-scan after iteration one, three and six when processing the measured data. At each iteration, the estimate is refined and the residuals are reduced in magnitude. We can note that the residuals for the

[2]The dark spots have negative amplitudes. Their magnitudes are approximately 1/5 of the main (bright) echoes.

measured data contain patterns that are more regular than the residuals for the simulated data. A probable cause of this is model errors. For a non-perfect model, the strongest contributions to the residual are expected where we have strong scatterers.



(a) First iteration.

(b) Third iteration.

(c) Sixth and last iteration.

Figure 3.5: Residual B-scans after a few of the iterations. The algorithm converged in six iterations

Computation time

To illustrate the computation times involved when using the algorithm, we here give the times required in the processing of the real data. A study showing how the computation time scales with the number of true scatterers and the image size can be found in [6].

The calculations were performed on a Pentium III with clock frequency of 2 GHz. The initialization of the matrix $\mathbf{B}$ took approximately 10 min. Note that this is a one-time calculation. When scanning a volume, several B-scans are collected under similar conditions and they all share the same matrix $\mathbf{B}$. The computation of $\mathbf{c}$, see Section 3.2.3, required 3.3 s.

The search required in total 3.5 s, which is approximately the same as the time for calculating $\mathbf{c}$. The time to perform each of the steps of the search algorithm is detailed in table 3.1. The five different steps in the method is explained in detail in Section 3.2.3.

It is clear that the updating part of the search dominates the search time. The single most time consuming operation is calculating the criterion $J(I \cup (m, n))$ for all potential candidates in step 1c. Calls to this operation consume a total of 1.94 seconds, or 55% of the total search time. In this case, approximately 7500 candidate points were evaluated in step 1c over the six

Step	Time [s]	Time [%]
1a	0.16	4.6
1b	0.2	5.7
1c	2.24	63.8
1d	0.7	19.9
2	0.21	6.0

Table 3.1: Computation time for different parts of the search algorithm.

iterations in total.

3.2.5 Conclusions and discussion

We have presented a new 2-d sparse deconvolution algorithm for deconvolution of ultrasonic B-scans. The simulation results show that the method under ideal conditions combines excellent resolution and noise suppression properties. The reason is that the available information is optimally used; we know that only a few significant scatterers are present and we therefore search for a only few significant scatterers. This also gives the advantage of a reduced computation time compared to the linear MMSE solution. Although an iterative search is performed, the computations in each iteration are far less demanding than those found in linear MMSE. For truly sparse images, the iterative search converges in only a few iterations.

One disadvantage with this method compared to the simpler SAFT methods is the small size of the images that can be processed and in the present form this method should mainly be seen as a complement to SAFT based system. Small ROIs can be chosen manually, or possibly automatically found using some simple method based on SAFT results, so that only small fractions of the data needs to be processed.

Note however that the overall time for processing the B-scan used in this paper, is approximately 7 s and this is less than the approximately 10 s that was used to collect the data. Thus it can for images of this size be implemented for real time processing.

Finally, a evaluation of the algorithm under less ideal premises than those given here should be performed in the future. Since we are approaching the limits of how much the linear model can be utilized, we need to take more robust methods under considerations. The challenge is to develop methods in which model uncertainties are taken into account and that, at the same time, can be implemented using computationally efficient methods.

3.3 Frequency domain SAFT

3.3.1 Background

Although advanced synthetic aperture focusing techniques (SAFT) implemented in frequency-domain have been widely used for many years in radar (synthetic aperture radar, SAR) and sonar (synthetic aperture sonar, SAS), they are relatively unused in nondestructive evaluation (NDE) of materials. Instead, simple time-domain SAFT has been applied in NDE for detection and characterization of defects in thick metal structures, especially in nuclear power plants [1], [2].

In SAFT the pulse-echo measurements made at a multitude of transmitter/receiver locations are combined to form a map of the ultrasonic reflectivity of the region of interest (ROI). The method takes advantage of both spatial and temporal correlations to enhance the resolution and the signal-to-noise ratio of the resultant images. However, the performance of SAFT in practical applications, especially in the near field, depends on the particular implementation of the algorithm as well as on the size of the transducer used in synthetic aperture [3]. Most SAFT implementations are based on a very simplified model of the imaging system used for developing radar and sonar applications. They can perform relatively well only provided that the assumptions that the ROI is located in the far field of the transducer (antenna) used for creating synthetic array where its specific diffraction effects can be neglected (point-like source assumption). While this assumption is generally valid for SAR and SAS, it is not always valid in ultrasonic imaging, especially in the high frequency NDE applications where the transducer can even be in contact with the inspected structure. Two principal problems may be encountered in such setup, firstly, transducer's diffraction effects may impair image quality, and secondly, sparse spatial sampling used for gathering ultrasonic data may yield artifacts in the resulting image. The analysis of SAFT performance for finite sized transducers should be of great interest for its potential users.

Accurate Fourier-based imaging techniques, developed recently in SAR, relate Fourier components of the measured SAR signal to the Fourier components of the target to be imaged. The origin of this approach can be found in the wave equation inversion theory, which is also known as wavefront reconstruction or holography [6], [13]. The basic principle behind wavefront reconstruction is the use of Fourier decomposition of the Green's function (also known as spherical phase function), which represents the impulse response of an imaging system [6].

In this chapter we propose a frequency-domain SAFT algorithm, which is a modified version of the wavenumber (ω-k) implementations known from SAR and SAS [6], [15], [16]. The algorithm is derived using a model developed in terms of wave equations. The model, which is valid in the far-field accounts for the beam-pattern of a finite sized transducer used in the synthetic aperture. Like in the ω-k implementations, the proposed algorithm employs the 2D FFT for transforming data between the time and frequency domains. A formal transform of the polar coordinate system, natural for ultrasonic transducers, to the Cartesian system suitable for the presentation of imaging results is used. However, since we intend to use the algorithm in the range interval where the point source assumption might not be valid we introduce a filter for compensation of the transducer lobe pattern. We compare the performance of the proposed algorithm to that of the standard time-domain (t-d) SAFT based on DAS operations with emphasis on the lateral resolution. The resolution and the side lobes of the algorithms are compared in the analysis and it is shown that the proposed algorithm consistently performs better than the

standard t-d SAFT.

The chapter is organized as follows; in the next section (3.3.2) we present theory starting from the basic relations defining the resolution and the sampling constrains that apply to the synthetic aperture setup. This presentation is followed by the models of the imaging system and the circular transducer that is used in the proposed algorithm. In the Section 3.3.3 results of the simulations performed for the t-d SAFT and the proposed algorithm are presented and their performance is compared. Finally, we show experimental results obtained in a simple setup with wire targets.

3.3.2 Theory

Resolution and spatial sampling in SAFT

Below, we will define spatial resolution of a synthetic aperture in monostatic strip-map mode (i.e., a linear equally sampled aperture without transducer beam-steering), which is an appropriate setup in many ultrasonic applications. There is an important difference between physical linear arrays and synthetic arrays which results in the synthetic aperture having a resolution finer than that corresponding to the real linear array of the same length focused in reception. Assuming that in a physical linear array the transmission results in an illumination of the ROI and the angle selectivity is performed in a receive beamformer, the differences in phase received by each element of the array contribute to its beam pattern. In the synthetic aperture, on the other hand, the same element radiates and receives signals and therefore the round-trip phase shift is effective in forming the resulting radiation pattern. As an important consequence of this fact the synthetic aperture has two times finer lateral (cross-range) resolution for the same aperture length, which can be expressed as

$$\delta_{3dB} \cong \frac{R\lambda}{2L_{eff}} \tag{3.15}$$

where, λ is the wavelength and δ_{3dB} is the effective half-power beamwidth of the synthetic aperture with length L_{eff} at a distance R, see [17]. The parameter L_{eff} denotes an effective aperture length, which is defined as the largest aperture length that, at the distance R contributes to the SAFT performance in terms of its lateral resolution. It is assumed that the signals received by all elements of a synthetic aperture are used efficiently if the L_{eff} is no longer than the half-power lobewidth of the transducer (element) used in the aperture. For a circular transducer with diameter d the half-power lobewidth at a distance R will be

$$L_{eff} = \cong \frac{R\lambda}{d} \tag{3.16}$$

Inserting this expression into (3.15) we get the fundamental relation defining lateral resolution of synthetic aperture

$$\delta_{3dB} \cong \frac{d}{2} \tag{3.17}$$

Thus the resolution obtained from a synthetic aperture is independent of frequency and range if the effective aperture length is used. Analysis of the support band of the SA signal in the cross-range direction leads to the condition for spatial sampling step

$$\Delta_x \leq \frac{R\lambda}{4L_{eff}} \tag{3.18}$$

Eq. (3.18) reduces to if the effective aperture defined by (3.16) is used for the shortest wavelength represented in the ultrasonic pulse (see [14], [16] for details). The relations presented above are correct for monochromatic wave and the targets located on the symmetry axis of the aperture.

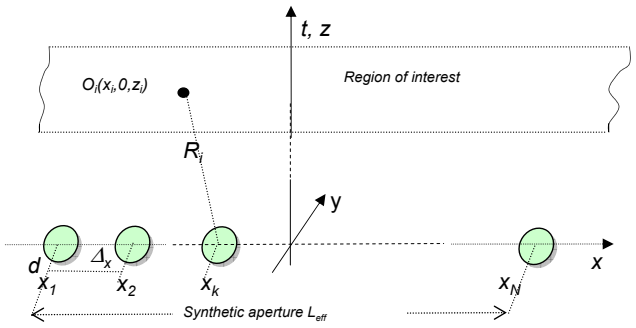


Figure 3.6: Geometry of synthetic aperture. At each position $\{x_1, x_2, \ldots, x_N\}$ of the synthetic aperture with length L_{eff} the transducer emits a pulse and receives en echo from a number of targets located in the range R_i at the points O_i.

Model of the Imaging System

In this section we present a general model of an imaging system that will enable deriving the imaging algorithm. The model, which is derived here for circular sources, can easily be generalized to other transducer geometries (for instance, rectangular array elements). Let us consider the synthetic aperture consisting of N transducer positions $\{x_1, x_2, \ldots, x_N\}$ in the setup shown in Figure 3.6. The aperture is created when a single transducer (or an array) performs N pulse-echo measurements at the positions $\{x_1, x_2, \ldots, x_N\}$, respectively. The transducer in its successive positions irradiates the region of interest located in the xz-plane. At each position the transducer (or the respective array element) emits an acoustic wave and receives an echo from an object defined by the reflectivity function $f(x, y, z)$ (e.g., including a finite number of point targets at points $O_i(x_i, 0, z_i)$). The successive transducer positions, located along the x-axis in the rectangular coordinate system xz are spaced with the pitch Δ_x. The transducer is excited by an electrical pulse $e_i(t)$ and it receives in each respective location $\{x_1, x_2, \ldots, x_N\}$ the electrical signals $s(x_k, t) \in \{s(x_1, t), s(x_2, t), K, s(x_n, t)\}$ that can be presented in the aggregated form referred to as B-scan image.

Our task is to perform imaging in a predefined region of interest located in front of the synthetic aperture at the symmetry plane xz, using the signal set $s(x_k, t)$, and in particular, to enhance the lateral resolution in this region. Following the notation used in SAR we will derive the imaging algorithm for continuous time and space model based on a number of measurements in the discrete points along the axis x. For each transducer position x_k the imaging model can be defined using the fundamental expression used for SAR imaging [14], [15]

$$s(x_k, t) = \int_x \int_z f(x, z)\delta\left(t - \frac{2}{c}\sqrt{z^2 + (x - x_k)^2}\right) dzdx \tag{3.19}$$

where x_k is transducer's position, $f(x, z)$ denotes object's reflectivity function, c is sound velocity and it is assumed that the transducer emits an impulse $\delta(\cdot)$.

Eq. (3.19) is valid assuming that the transducer diameter d is small compared to the wavelength λ, in other words the transducer can be regarded as a point source, that is, its specific diffraction effects can be neglected. This implies that the response of the single point target takes the form of a hyperbola in the B-scan and the assumptions required for the standard SAFT based on the DAS operations are fulfilled. Below, we will derive a more realistic transducer model to the eq. (3.19) including its far-field beam pattern and its electrical frequency response.

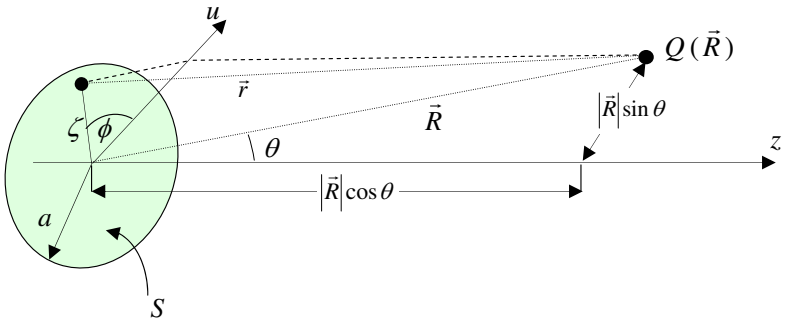


Figure 3.7: Geometry for calculating the far-field response of a circular transducer.

Transducer Model

Here, we will consider the case when a finite sized transducer is used for the measurements and its diffraction pattern has to be taken into account. However, our analysis will be confined to the far-field zone where the Fraunhofer approximation can be used. We start by deriving an expression defining the signals $s(x_k, t)$ for a single circular transducer. This expression will be then used for the derivation of the synthetic aperture scheme in frequency-domain. Let us consider a circular piston transducer of radius a in the polar coordinate system shown in Figure 3.7. Since the same transducer is used for all the measurements we introduce a separate coordinate system with its origin at the transducer position x_k. The incident pressure $P_{inc}(\vec{R}, w)$ at a general off-axis point $Q(\vec{R})$ in far-field can be found from

$$P_{inc}(\vec{R}, \omega) = -j\omega\rho v_0 a^2 \frac{e^{-jkR}}{R} \frac{J_1(ka\sin\theta)}{ka\sin\theta} \tag{3.20}$$

where J_1 is the first order Bessel function, ρ denotes the specific medium density, v_0 represents the normal particle velocity at the transducer's surface, k is the wave number, R is the distance $|\vec{R}|$ and $a = d/2$ is the transducer radius [18], [19]. Eq. (3.20) consists of three factors: the first is a frequency dependent coefficient, the second represents a spherical wave, while the third term, often referred to as the *jinc* function defines the angular dependence of the amplitude of this spherical wave. The *jinc* function indicates the lobe structure of the beam in far-field characterized by the main central lobe and a number of side lobes. It is worth noting that the assumptions used by standard SAFT algorithms are valid only for small values of ka when the main lobe is wide and the amplitude of the spherical wave varies very little with the angle θ.

The point target with an elementary surface s_e located in the point $Q(\vec{R})$ will scatter back the incident pressure wave and the pressure integrated over the transducer surface $\bar{P}_r(\vec{R}, \omega)$ is proportional to the convolution (the multiplication in frequency domain) of the incident pressure $P_{inc}(\vec{R}, w)$ and the pressure reflected from the target surface $P_{ref}(\vec{R}, w)$ (see [20] for details)

$$P_r(\vec{R}, \omega) = P_{inc}(\vec{R}, \omega) P_{ref}(\vec{R}, \omega) = -\frac{s_e}{2}\omega^2(\rho v_0 a^2)^2 \frac{e^{-jkR}}{R^2} \mathrm{jinc}^2(ka\sin\theta) \tag{3.21}$$

Note that (3.21) is valid for a monostatic configuration if the transmission and reception of the transducer are separated in time and the medium is isotropic. If the transducer is modeled as a linear electro-mechanical system with electrical frequency responses in transmission and reception, $H_{et}(\omega)$ and $H_{er}(\omega)$, respectively, the received signal will be a convolution of the incident pressure and transducer's electrical characteristics

$$S(\vec{R},\omega) = H_{et}(\omega)H_{er}(\omega)\bar{P}_r(\vec{R},\omega) = \bar{P}_r(\vec{R},\omega)H_e(\omega) \tag{3.22}$$

where the joint electrical transducer frequency response is $H_e(\omega)$ and it is assumed that the pulse $u(t) = \delta(t)$ is used to excite the transducer. Finally, by inserting (3.21) into (3.22) the received signal is

$$S(\vec{R},\omega) = \frac{s_e}{2}(\rho v_0 a^2)^2 \frac{e^{-jkR}}{R^2}\text{jinc}^2(ka\sin\theta)\omega^2 H_e(\omega) \tag{3.23}$$

(the minus sign was omitted, see [19] for details). The above expression is a product of four factors: the first is a constant, the second represents a spherical wave that have propagated to the target and back, the third defines the diffraction effect of the transducer, and finally, the fourth denotes the second derivative of the transducer's electro-mechanical transfer function.

An important consequence of (3.23) is that the main lobe width for a finite sized transducer is limited by the *jinc* function, which as mentioned above, limits the maximum length of the synthetic aperture. Verifying the above condition for a given transducer diameter d one should keep in mind that the *jinc* has its -3dB lobewidth $\theta_{3dB} = 3.232/ka \cong \lambda/d$. If θ_{3dB} is small, the maximum length of a synthetic aperture for a given distance R will be $L_{max} = R\theta_{3dB} = R\lambda/d$.

Below, we will confine us to the targets located at the symmetry plane of the synthetic aperture, i.e., we will only consider points $O(x,0,z)$ in the ROI, for those $R = \sqrt{x^2+y^2}$. Then, assuming that distance compensation $(1/R)$ is performed in the receiver, we obtain after rearranging (eq:SRw2) the response of the transducer located at the position x_k to a single scatterer at the point $O(x,0,z)$

$$S(\vec{R},\omega) = S(x_k,\omega) = \frac{s_e}{2}(\rho v_0 a^2)^2 e^{-jk\sqrt{z^2+(x-x_k)^2}}A(\omega,k_x)\omega^2 H_e(\omega) \tag{3.24}$$

where $A(\omega,k_x) = jinc^2(ka\sin\theta) = jinc^2(k_x a)$.

To obtain a signal received by the transducer located at that point, given the real image of the point scatterers in front of the transducer $f(x,z)$ we have to integrate the transducer's response over the whole area

$$S(x_k,\omega) = \alpha A(k_x)\omega^2 H_e(\omega)\int_x\int_z f(x,z)e^{-j2k\sqrt{z^2+(x-x_k)^2}}dzdx \tag{3.25}$$

where α is a constant coefficient.

To perform synthetic aperture imaging in frequency domain we need the 2D Fourier transform of the signal $s(\vec{R},t)$, which means that (3.25) has to be further transformed from the space coordinate to spatial frequency (referred to as Doppler wavenumber in sonar). It can be shown that using the principle of stationary phase [6], [7], the 1D Fourier transform of the above equation along the x-direction can be expressed as

$$S(x_k,\omega) = \alpha A(k_x)\omega^2 H_e(\omega)\int_x\int_z f(x,z)e^{-j\sqrt{4k^2-k_x^2}z}e^{-jk_x x}dzdx \tag{3.26}$$

If we now introduce a new set of coordinates, defined by

$$Sk_z(k_x) = \sqrt{4k^2-k_x^2} \tag{3.27}$$

and denote $H_{ed} = \omega^2 H_e(\omega)$, we can finally express $S(k_x,\omega)$ as

$$S(x_k, \omega) = \alpha A(k_x) H_{ed}(\omega) F(k_x, k_z) \tag{3.28}$$

Note that now the coordinates of $S(k_x, \omega)$ and $A(k_x)H_{ed}(\omega)$ and those of $F(k_x, k_z)$ are different. The coordinate transformation defined by (3.27), which is known as Stolt mapping will be denoted as $\mathcal{S}\{\cdot\}$ in the sequel.

Imaging Algorithm

The *wavenumber* reconstruction algorithm presented in [7] and [15] takes the form of a spatio-temporal matched filter and the Stolt coordinate transform that can be summarized by

$$\hat{F}(k_x, k_z) = \mathcal{S}^{-1}\left\{ e^{j(\sqrt{4k^2 - k_x^2} - 2k) r_o} A^*(k_x) H_{ed}^*(\omega) S(k_x, \omega) \right\} \tag{3.29}$$

where r_0 is the distance to the middle of ROI and asterisk denotes a complex conjugate. The main function of the algorithm (3.29) is compensating the phase shift introduced by the imaging system (the $e^{[\cdot]}$ term) and transforming the coordinate system (the $\mathcal{S}^{-1}\{\cdot\}$transform). The term $H_{ed}^*(\omega)$ in (3.29) performs compression of the electrical impulse, and the term is intended to compensate the effect of angular dependence of the transducer amplitude. Since the latter is not performed satisfactory by the matched filter (3.29) a modified version using a Wiener filter for the beam pattern compensation is proposed here

$$\hat{F}(k_x, k_z) = \mathcal{S}^{-1}\left\{ e^{j(\sqrt{4k^2 - k_x^2} - 2k) r_o} \frac{A^*(k_x)}{A(k_x)A^*(k_x) + \varepsilon} H_{ed}^*(\omega) S(k_x, \omega) \right\} \tag{3.30}$$

The small constant ε should limit the filter output outside the support band of $A(k_x)$. There are at least two reasons why the electrical frequency response $H_{ed}(\omega)$ is not included in the Wiener filter (3.29), first, it is measured separately with a limited accuracy why the $A^*(k_x)$ is calculated analytically. Second, in practical situations the $H_{ed}(\omega)$ will also include the space-variant low-pass effect due to the transducer spatial impulse response. Thus the filter used for the compensation of transducer's frequency response has to be robust enough to perform well in the presence of model errors. Both practical experience and the theoretical analysis presented in [21] show that in such situations the parameter-free matched filter (complex conjugate of the $H_{ed}(\omega)$) is much more robust than the Wiener filter.

The ω-k algorithm (referred to as ω-k SAFT in the sequel) consists of three main steps:

1. 2D Fourier transform of the acquired data $s(x_k, t) \rightarrow S(k_x, \omega)$,
2. Filtering and change of variables using (3.29), or (3.30) and
3. Inverse 2D Fourier transform $\hat{F}(k_x, k_z) \rightarrow \hat{f}(x, z)$.

The ω-k SAFT can be easily implemented and fast executed using the existing FFT routines. It compensates for the transducer's diffraction distortion as well as its electromechanical characteristics, which is a great advantage in contrast to the t-d SAFT based on DAS operations. Note however, that the above applied transducer model is valid in far-field only and in the near-field the low-pass filtering effect due to the finite lengths of transducer's SIRs must also be compensated for, [3]. Although the ω-k SAFT was derived above for the monostatic configuration the extension to a bistatic or multistatic setup is quite straightforward and consists in modifying the imaging system model starting from eq. (3.21) where different sound paths are to be accounted for. This finally results in a modified phase compensation term in eqs (3.29)

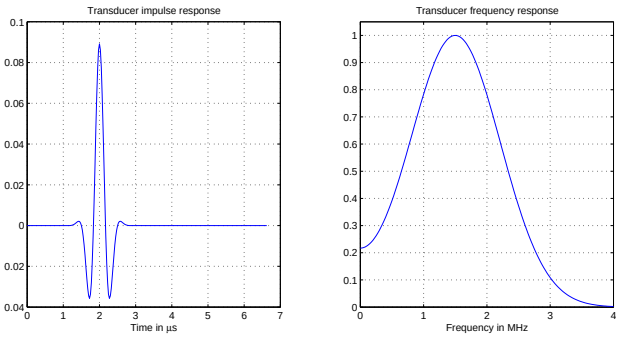


Figure 3.8: Excitation pulse used in simulations and its Fourier transform.

and (3.30). The main disadvantage of the ω-k algorithms is the need to interpolate from one 2D sampling grid to another grid defined by the Stolt mapping. Indeed, by virtue of physics ultrasonic data acquired by a transducer correspond to a curved surface while the real image traditionally is reconstructed in a rectangular grid. Below, we illustrate the performance of the ω-k SAFT using simulations and we compare it to that of the standard t-d SAFT.

3.3.3 Simulations

The simulations were performed for the synthetic aperture created from the measurements made by circular transducers with various diameters d for a point target located in water at different distances z_t from the aperture. Time domain simulations were used and the spatial impulse responses of circular transducers in the simulations were calculated from analytical solutions [22]. Standard pixel-driven t-d SAFT implementation was used in the simulations. In this implementation the theoretical hyperbolas (a binary pattern) corresponding to the respective ranges are shifted across the processed RF 2D data and the values of pixels where the hyperbola crossed the data are summed. A linear interpolation was used to compensate the effects of discretization. This implementation can be thought of as a spatially-variant matched filter employing the theoretical response of the point target received by a point transducer with an unlimited bandwidth. The output of the matched filter applied to a real RF ultrasonic signal, due to the oscillatory character of the transducer response can take both positive and negative values.

The ω-k SAFT was implemented according to (3.30), which means that the beam pattern compensation was used in all results presented below, except Figure 3.13 comparing the resolution of different algorithms. Contrary to the t-d SAFT, which processed RF signals, the ω-k SAFT was implemented on complex-valued quadrature demodulated signals. To facilitate comparisons with the ω-k SAFT the envelopes (calculated using Hilbert transform) of the t-d SAFT results are presented in figures below, except Figure 3.9 and Figure 3.15 where the rectified amplitudes are presented. The lateral resolution of both algorithms is compared using projection of the cross-range beam profiles on the transducer plane (the xy plane). The range resolution is illustrated by the envelopes of the center A-scans at x=0 in the B-scans.

The apertures used for the t-d SAFT included the number of elements calculated for each target distance and each transducer diameter according to (3.16), given a constant pitch $\Delta_x \leq \frac{d}{4}$. The transducers were excited by the broadband pulse shown in Figure 3.8 with its center frequency at 1.5 MHz.

The first simulation shown in Figure 3.9 and 3.10 was performed for a circular transducer with diameter 4 mm and a point target located at 40 mm. The aperture used for t-d SAFT

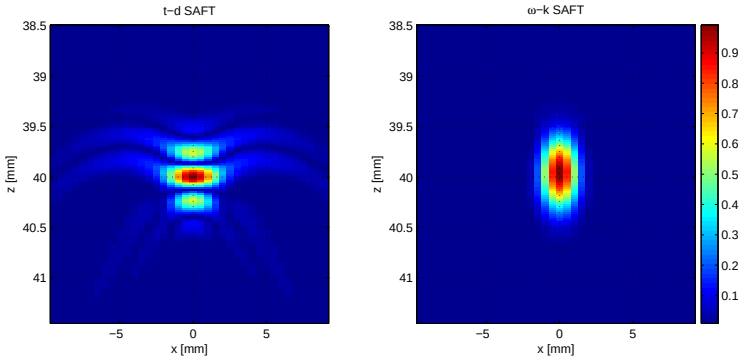


Figure 3.9: Plots of the normalized ampitudes obtained for the point target located at distance 40 mm from aperture after processing with t-d SAFT (left) and the ω-k SAFT (right). Note that the absolute value of the t-d SAFT result is presented.

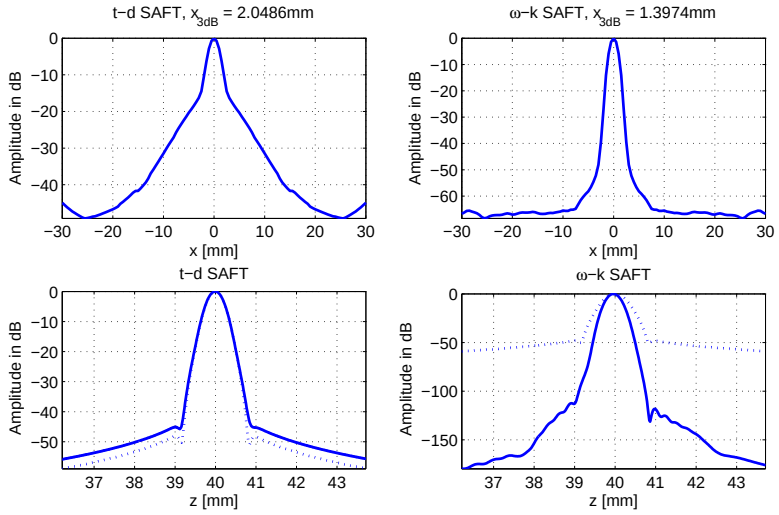


Figure 3.10: Beam projections obtained from B-scans shown in Figure 3.9 in dB. Upper: lateral profile obtained from the t-d SAFT (left) and the ω-k SAFT (right). Note the values of 3dB resolution (beam width) printed above the respective panels. Lower: envelope of the center A-scan at x=0 in the B-scan obtained from the t-d SAFT (left) and the ω-k SAFT (right). The dotted line represents the enevelope of the original A-scan (raw data).

consisted of 20 elements, spatial sampling was 0.5 mm, and the proposed ω-k SAFT processed the ROI $x \in [-30, 30], z \in [25, 55]$ (only a small center part of the ROI is shown in Figure 3.9). From Figure 3.9 and 3.10 it can be seen that the lateral resolution of ω-k SAFT is much better that that of t-d SAFT (the -3dB beam profile widths x_{3dB} evaluated using interpolation with second order polynomial were respectively, 1.4 mm and 2.0 mm). The ω-k SAFT profile has its side lobes at the level less than -60 dB while the t-d SAFT profile is characterized by broad side artifacts seen in Figure 3.9 and clearly pronounced as a broad lobe in Figure 3.10 (upper part).

The temporal resolution of both algorithms can be evaluated from the lower part of Figure 3.10 where envelopes of A-scans at $x = 0$ in the respective B-scans are plotted for t-d SAFT and the ω-k SAFT (the corresponding A-scan profile from the raw data is also plotted as a reference). It is apparent that the ω-k SAFT results in a much better temporal resolution than that obtained from t-d SAFT (the respective x_{3dB} is smaller). From Figure 3.10 can be seen that the range profile obtained from t-d SAFT has essentially the same width as the raw data, while the ω-k SAFT yields much smaller profile decreasing to approximately -120 dB.

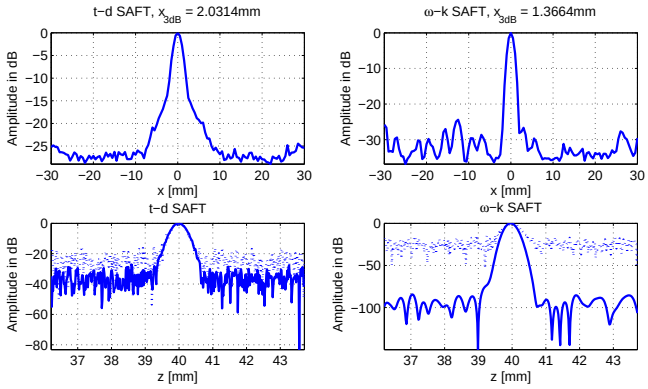


Figure 3.11: Beam projections obtained after processing simulated data from the same setup as that in Figure 3.10 but corrupted by Gaussian noise. Upper: lateral profile obtained from the t-d SAFT (left) and the ω-k SAFT (right). Note printed values of x_{3dB} (3 dB beam width) as a measure of resolution. Lower envelope of the center A-scan at $x = 0$ in the B-scan obtained from the t-d SAFT (left) and the ω-k SAFT (right), dotted line represents the original B-scan profile (raw data).

To evaluate robustness of both algorithms in the presence of measurement noise the same experiment was repeated again but white zero-mean Gaussian noise was added to the (raw) simulated data. The profiles presented in Figure 3.11 show that the lateral profile width x_{3dB} is essentially unchanged and the ratio of the main lobe amplitude to the noise level is similar for both algorithms, approx. 25 dB (for the corresponding noise ratio approx. 20 dB in the raw data). The range profile for the ω-k SAFT is much better than that for t-d SAFT. The maximum signal to noise ratio is approx. 30 dB higher.

The lateral resolution of both algorithms for different transducer sizes (but the same electrical impulse response shown in Figure 3.8) for a target located at a distance 100 mm can be compared in Figure 3.12 (the -3dB profile width of the raw B-scan is plotted as a reference). Figure 3.12 can be seen that the resolution of both algorithms decreases proportionally (the x_{3dB} increases) to the transducer diameter. However, the ω-k SAFT has the superior resolution, which is better than that predicted by (3.17).

For large transducer diameters, $d > 10$ mm for the distance 100 mm shown in Figure 3.12, the application of the synthetic aperture does not yield improvements in resolution. It is apparent that for large transducers, as the angular information in the target response decreases, the use of t-d SAFT algorithms does not result in significant improvement of resolution. In other words, the transducer lobe becomes so narrow that processing consecutive A-scans does not improve the lateral resolution. Note that this is a direct proof of the general rule according to which synthetic aperture can be effective only if the point source assumption is fulfilled.

In Figure 3.13, the resolution obtained with t-d SAFT and the ω-k SAFT are presented as a function of target distance for transducer diameter d = 4 mm. The results obtained for the ω-k SAFT without aperture compensation, eq. (3.29) are plotted besides those obtained for the proposed algorithm with Wiener filter, eq. (3.30). All three algorithms keep constant resolution as predicted by 3.17. However, while the resolution of t-d SAFT is approx. equal to $d/2$ the resolution of the ω-k SAFT is higher (smaller x_{3dB}). The difference between t-d SAFT and the ω-k SAFT implementation is approximately 30% (taking t-d SAFT's resolution as a reference). The effect of beam pattern compensation using Wiener filter (eq. 3.30) is clearly pronounced, the respective x_{3dB} of the ω-k SAFT with compensation is approximately 5% smaller for all distances.

Summarizing, the simulations have shown that the ω-k SAFT offers a clear performance

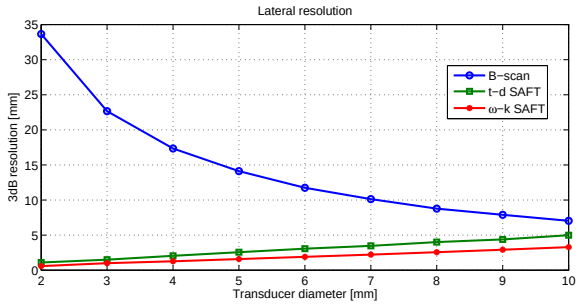


Figure 3.12: Lateral resolution (the x_{3dB}) as a function of transducer diameter for the t-d SAFT and the ω-k SAFT for target distance 100 mm. The 3dB profile width of the ultrasonic data (B-scan) is plotted for reference.

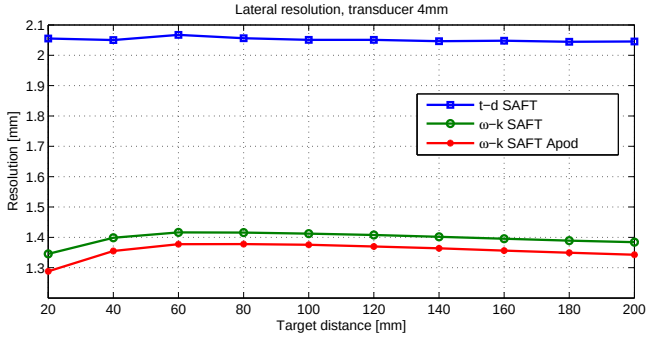


Figure 3.13: Lateral resolution in terms of the profile width x_{3dB} as a function of target distance for the t-d SAFT and the ω-k SAFT, respectively without and with aperture compensation for transducer diameter d = 4mm.

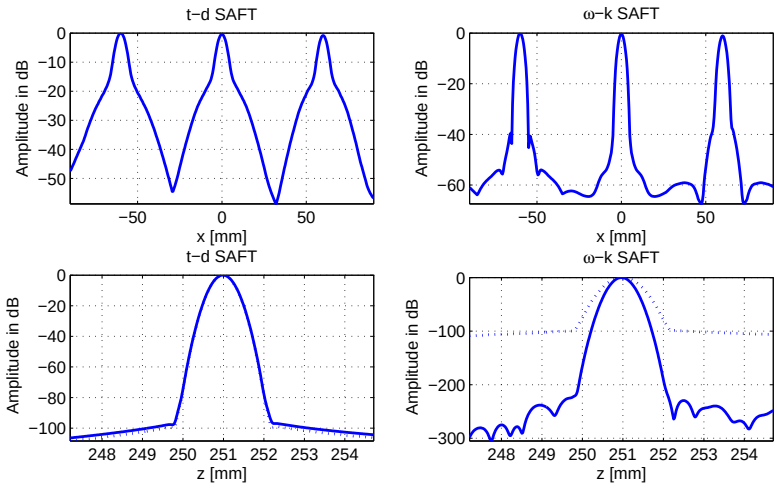


Figure 3.14: Simulation results obtained in the experimental setup for 3 targets spaced 30 mm from each other (the middle target at the distance 251mm from SA). Upper panel: lateral profile obtained using t-d SAFT (left) and the ω-k SAFT with lobe compensation (right). Lower panel: range profiles obtained for the middle target using the t-d SAFT (left) and the ω-k SAFT (right), dotted line represents the original B-scan profile (raw data).

improvement comparing with the t-d SAFT in terms of resolution and lower side lobes. An additional improvement is obtained by introducing beam pattern compensation in the ω-k SAFT (eq. 3.30).

3.3.4 Experiment

To verify the performance of the compared algorithms, a simple experiment was carried out in a water tank. The measurements were performed using a planar 0.375" 2.25 MHz immersion transducer V325-SU from Panametrics. Three steel wires of diam. 0.2 mm immersed in water, respectively at the distance $z_t = 221$, 251 and 280 mm from the aperture were used as targets. The transducer was moved in 1 mm steps and the ultrasonic data was acquired with sampling frequency 80 MHz and digitized using an 8-bit AD converter.

The experimental setup was simulated using the software tools used for the simulations presented in the former section. The simulation results are summarized in Figure 3.14 where lateral profiles are presented in the upper panels and the temporal resolution is illustrated by the graphs in the lower panel. The lateral resolution x_{3dB} estimated using interpolation of profiles with the second order polynomial was respectively, 4.9 mm for t-d SAFT and 3.3 mm for the ω-k SAFT with lobe compensation. The resolution was similar for all 3 targets spaced 60 mm from each other. The side lobe level was respectively -20 dB for t-d SAFT and -40 dB for the ω-k SAFT. Also the temporal resolution of the ω-k SAFT was much better that that of t-d SAFT (cf. lower panel of Figure 3.14).

The results obtained from the ultrasonic measurements performed in the simulated setup are presented in Figure 3.15. The B-scans and lateral profiles are shown respectively for raw data (upper panel), the t-d SAFT processed data (middle panel) and the data processed using the ω-k SAFT with lobe amplitude compensation (lower panel). Amplitudes of the raw data were distance-compensated after the acquisition. Both t-d SAFT and the ω-k SAFT yield clear improvement of the lateral resolution, the values estimated in the same way as in the simulations are respectively, 4.6 mm for SAFT and 3.6 mm for the ω-k SAFT. Both values are very close to those obtained in simulations.The t-d SAFT has well pronounced sidelobes both at the B-scan

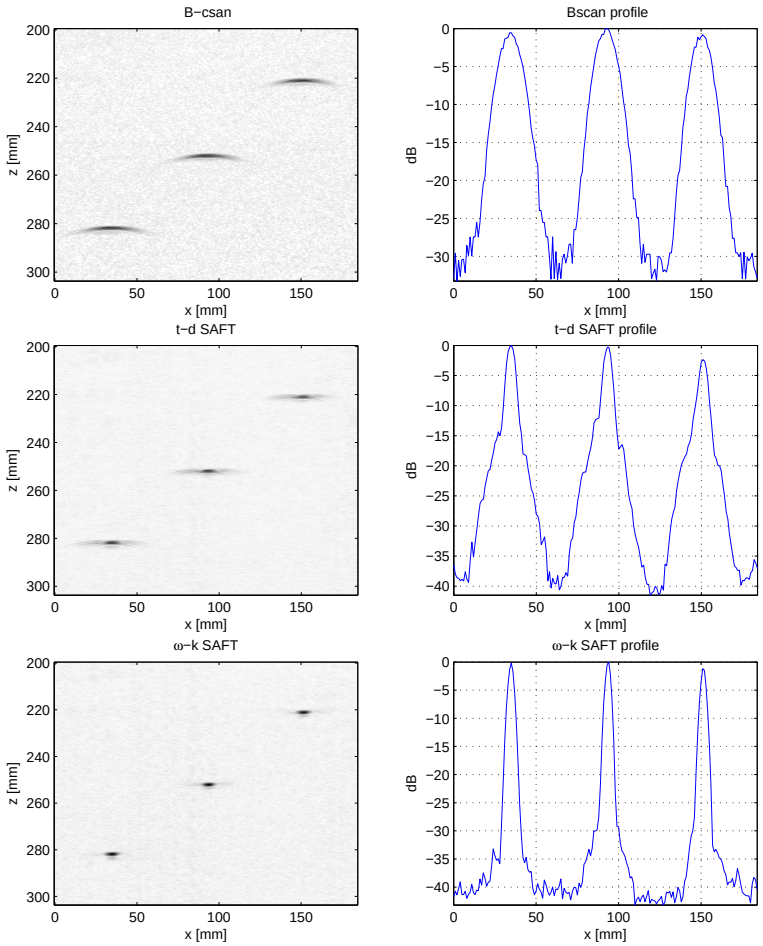


Figure 3.15: Results obtained in the experimental setup for 3 immersed targets spaced 30 mm in range (the middle target at the distance 251mm from the SA). Upper panel: B-scan (left) and lateral profile obtained from raw data (right). Middle panel: B-scan (left) ant the lateral profile obtained using the t-d SAFT (right). Lower panel: B-scan (left) and the lateral profiles obtained for the ω-k SAFT with lobe compensation (right). Note that the B-scans of raw data and the t-d SAFT result are rectified. For better presentaion of the background noise level pixel amplitude in the B-scan images was compressed using the square root function.

and at the cross-range profile at the level of -15 dB. The sidelobes of the ω-k SAFT cannot be seen in the B-scan with linear amplitude coding but the corresponding value read out of the profile is approx. -32 dB. The amplitudes of all targets are well preserved both by the t-d SAFT and by the ω-k SAFT.

The origin of the sidelobes present in the t-d SAFT imaging can be understood if the responses in the t-d SAFT imaging result, shown in Figure 3.9 and 3.15 are analyzed closer. It can be seen that the main lobe corresponding to each target has a kind of "wings" at its both sides. The "wings" originate form the simplified implementation of t-d SAFT which does not account for the transducer electrical impulse response. Indeed, when a single pixel thick hyperbola is shifted across the processed B-scan (see the description of the t-d SAFT implementation in the introduction to Section 3.3.3) a number of overlapping points will be detected due to the finite width of the transducer response and this effect is particularly well pronounced in the neighborhood of the main peak corresponding to the target position. The ω-k SAFT does not have this drawback since the transducer's electrical frequency response is included in the model of the imaging setup.

In summary, the experimental result is in good agreement with the simulated one and it confirms the superior performance of the proposed ω-k SAFT.

3.3.5 Conclusions

The ω-k SAFT for synthetic aperture imaging proposed in the paper is based on the frequency domain model of the imaging system. The algorithm performs the 2D FFT transform of the measured ultrasonic data followed by the 2D matched filter and the Stolt coordinate transform. The result is transformed back to time-space domain using inverse FFT. Wiener filter based on the far-field model of transducer beam pattern is proposed for the compensation of the main lobe amplitude variation.

The simulated results revealed that the ω-k SAFT offers a clear performance improvement comparing with the standard time-domain SAFT in terms of the improved both range resolution and lateral resolution as well as lower side lobes in the images of point targets. The lateral resolution of the ω-k SAFT in a wide range of target distances is approximately 30% better that the theoretical limit equal to half diameter of the transducer used in the synthetic aperture.

The transducer beam pattern compensation based on the Wiener filter concept yields a clear, constant for all target distances, improvement of the lateral resolution. The monostatic configuration presented in the paper can easily be replaced by a bistatic or multistatic setup. Similarly, the transducer model, which was derived here for the circular sources, can easily be generalized to other transducer geometries, for instance, rectangular array elements. When doing so one should bear in mind, however, that SAFT requires half of the pitch used in conventional array systems.

3.4 Fast computation of a symmetric matrix product

One of the most time consuming steps in the both the ESAFT algorithm and the sparse deconvolution method presented above in this chapter is the multiplication of the spatial impulse response matrix $\mathbf{P}$ with its transpose: $\mathbf{B} = \mathbf{P}^T\mathbf{P}$. $\mathbf{P}$ contains a great deal of symmetry that arises from the linear imaging model used in these methods. This symmetry could be exploited to speed up the calculation of $\mathbf{B}$. In this section a simple scheme to exploit available symmetry in $\mathbf{P}$ is presented. Details on how $\mathbf{P}$ is constructed and structured is given in section 3.4.1, the algorithm itself is presented in section 3.4.2. In section 3.4.3 computation times for several data sets of different sizes are presented to evaluate the performance of the suggested method.

3.4.1 Model of the imaging system

Both the ESAFT method, presented in [12], and the sparse 2D deconvolution method presented in section 3.2 use a similar linear model of the imaging system. Consider the measurement setup shown in Figure 3.1. The region of interest (ROI) that is imaged is denoted by $\mathbf{O}$. The ROI that is divided into $M \times N$ image elements is located in the xz-plane with $y = 0$. Every element $o(m, n)$ in the image $\mathbf{O}$ is a scalar representing the scattering strength of a target in the position (m, n). Measurements from L positions of a transducer moving parallel to the x-axis are available in discrete-time form. The measurement set consists of L A-scans, each containing K samples. It is common, but not necessary, that $K = M$ and $L = N$ so that there is one image element for each sample in the measurements. Another common sampling of the image is choosing $N = \eta L$ and $M = K/\xi$ where η and ξ are integer constants. That is: For each transducer position, estimate η columns of image elements and for each ξ samples in the measured data, estimate one row of image elements. Let the $K \times 1$ vector $\mathbf{y}_l$ denote the noise-free A-scan response at position T_l, which can be expressed as a sum of echoes from all scatterers in the ROI. Let $\mathbf{p}_l(m, n)$ denote the $K \times 1$ vector that consist of the combined time discreet electrical and spatial impulse response of a scatterer at position (m, n) and a transducer at position l. Calculating the SIRs of a transducer naturally requires information about its geometry and focusing, but even for complicated geometries the SIRs can be calculated using numerical methods.

By neglecting multiple scattering effects, an A-scan can be modeled as a superposition of the contributions from all scatterers in the ROI. Using the notation developed above, the A-scan can be written as

$$\mathbf{y}_l = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \mathbf{p}_l(m, n) o(m, n). \tag{3.31}$$

The L vectors $\mathbf{y}_0, ..., \mathbf{y}_{L-1}$ can now be stacked on each other to form a $KL \times 1$ vector, $\mathbf{y}$. In a similar way a $MN \times 1$ vector, $\mathbf{o}$ is formed by stacking all elements in $\mathbf{O}$. Form the size $K \times M$ matrices

$$\mathbf{P}_{(l,n)} = (\mathbf{p}_l(1, n) \dots \mathbf{p}_l(M, n)) \tag{3.32}$$

for all combinations of transducer positions l and columns in $\mathbf{O}$. Unless the transducer geometry or focusing changes from one measurement position to the next, it will always hold that $\mathbf{P}_{(l,n)} = \mathbf{P}_{(l+a,n+a)}$, where a is some integer constant. This means that we can use the short notation $\mathbf{P}_{\tilde{n}}$ where $\tilde{n} = l - n$. Now the complete SIR matrix $\mathbf{P}$ can be constructed. If the simple relation

$L = N$ holds, $\mathbf{P}$ can be written as:

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_0 & \mathbf{P}_1 & \dots & \mathbf{P}_{N-1} \\ \mathbf{P}_{-1} & \mathbf{P}_0 & \dots & \mathbf{P}_{N-2} \\ \vdots & & & \vdots \\ \mathbf{P}_{-N+1} & \mathbf{P}_{-N+2} & \dots & \mathbf{P}_0 \end{pmatrix}, \tag{3.33}$$

If $N = \eta L$, then only every η block row in equation (3.33) should be included. We will assume that $L = N$ in the rest of this section. This does not imply a loss of generality, but it will simplify the notation considerably.

The matrix $\mathbf{P}$ contains all the combined electrical and spatial impulse responses for all combinations of points (m, n) in the image and all transducer positions l. Using this matrix, it is possible to write the model in the compact notation

$$\mathbf{y} = \mathbf{P}\mathbf{o} + \mathbf{e}, \tag{3.34}$$

where $\mathbf{e}$ is additive measurement noise. See [22, 24] for details concerning the impulse responses, the structure of $\mathbf{P}$ and other details of this model.

In both the ESAFT algorithm and in the sparse 2D deconvolution algorithm the matrix $\mathbf{B} = \mathbf{P}^T\mathbf{P}$ needs to be computed. Direct calculation of a matrix product of this size require $(MN)^2(KL)$ multiplications. Since MN and KL typically are in the range 3.000 - 15.000 this operation can take quite some execution time to complete, it is not uncommon for this step to take in the order of up to 15 minutes.

3.4.2 Calculation of B

Please note that that in equation (3.33) there is only a few blocks $\mathbf{P}_{\tilde{n}}$ that are repeated in several places in $\mathbf{P}$. This structure in $\mathbf{P}$ can be exploited in the calculation of $\mathbf{B} = \mathbf{P}^T\mathbf{P}$. The more symmetric the imaging system is, the more symmetries will be available in $\mathbf{P}$. In many applications, the spatial impulse responses are identical for observation points equally far off the central axis of the transducer but in different directions and $\mathbf{P}_{\tilde{n}} = \mathbf{P}_{-\tilde{n}}$.

Consider the matrix $\mathbf{B}$. It can be divided into $N \times N$ blocks, each $M \times M$ elements in size. Each block can be written as a sum of L products of blocks from two block columns of $\mathbf{P}$:

$$\mathbf{B}_{i,j} = \sum_{q=i,r=j}^{q=i-L,r=j-L} P_q^T P_r \tag{3.35}$$

Since each block $\mathbf{B}_{i,j}$ requires L block products, there is a total of N^2L products to be calculated. But since there are only $2N-1$ different blocks in $\mathbf{P}$ there is less than $4N^2$ different block products $P_n^T P_l$ possible. Also, since $P_n^T P_l = P_l^T P_n$ only a total of $2N^2$ different block matrix products actually have to be calculated, the rest can be found from a simple transpose. In the common case that $\mathbf{P}_{\tilde{n}} = \mathbf{P}_{-\tilde{n}}$, mentioned above, then the total number of different blocks is only N, thus reducing the number of unique block products to less than N^2. The following scheme exploits this symmetry in the calculations:

For each valid combination of q and r:

(a) Calculate $\mathbf{B}_{qr} = P_q^T P_r$.

(b) For each block element $\mathbf{B}_{i,j}$ in $\mathbf{B}$

If the newly calculated $\mathbf{B}_{qr}$ is a term in the sum making up this element, let $\mathbf{B}_{i,j} = \mathbf{B}_{i,j} + \mathbf{B}_{qr}$.

(c) Calculate $\mathbf{B}_{rq} = \mathbf{B}_{qr}^T$

(d) For each block element $\mathbf{B}_{i,j}$ in $\mathbf{B}$

If the newly calculated $\mathbf{B}_{rq}$ is a term in the sum making up this element, let $\mathbf{B}_{i,j} = \mathbf{B}_{i,j} + \mathbf{B}_{rq}$.

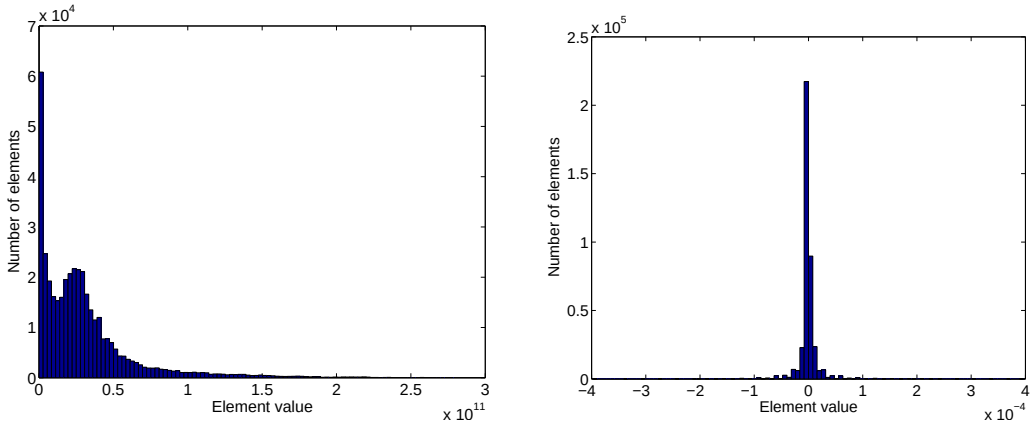
In this algorithm, each matrix multiplication is only calculated once, then added to all the blocks $\mathbf{B}_{i,j}$ where it applies. Each block product $P_q^T P_r$ requires M^2K operations, so this scheme requires M^2N^2K operation, reducing the total amount of multiplications with a factor L.

3.4.3 Simulations

The presented algorithm was tested for several different sizes of the image. Only the simplest case where $K = M$, $N = L$ and a transducer symmetric around the z-axis is considered. All calculations were performed on a Pentium III with clock frequency of 2 GHz and 4 GB of RAM memory. $\mathbf{B} = \mathbf{P}^T\mathbf{P}$ was calculated both directly in MATLAB and with the method presented here, implemented also in MATLAB.

Accuracy

To verify that the new method finds the correct value, $\mathbf{B}$ was calculated using both methods for an example imaging geometry. An histogram of both $\mathbf{B}$ and the difference between the two methods is shown in figure 3.16.



(a) Histogram for $\mathbf{B}$ calculated with the standard Matlab call $\mathbf{P}' * \mathbf{P}$

(b) Histogram for the difference between $\mathbf{B}$ calculated with the standard Matlab call and the new method.

Figure 3.16: Histograms for $\mathbf{B}$ and the error.

As is shown in the histograms, the difference between the two methods are very small, and are due to problems with numerical precision.

Calculation time

In table 3.2, the computation time for calculation $\mathbf{B}$ using both methods for some different image sizes is presented. It is evident that the new method requires far less computations than the

Size	N	M	L	K	Comp time (Matlab)	Comp time (New)
1260×1260	21	60	21	60	1.25s	0.65s
2790×2790	31	90	31	90	13.2s	2.79s
4920×4920	41	120	41	120	75.5s	8.76s
7650×7650	51	150	51	150	283.6s	25.3s
10980×10980	61	180	61	180	906.45s	61.8s

Table 3.2: Computation times as function of size of $\mathbf{B}$.

direct approach. Computation times also grow with image size as expected from section 3.4.2. The computation time is not reduced as much as a factor L, as there is some overhead for repeatedly accessing memory for different sections of $\mathbf{B}$.

Conclusions

A method to reduce the computational load when performing one of the steps in the ESAFT algorithm is presented. The method exploits known structures and symmetries in the SIR matrix $\mathbf{P}$. Simulations show a substantially decreased computation time for most common symmetries found in the SIR matrix $\mathbf{P}$.

Bibliography

[1] J. Shen and E.S. Ebbini. A new coded-excitation ultrasound imaging system–part I: Basic principles. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 43(1):131–140, January 1996.

[2] F. Lingvall, T. Olofsson, and T. Stepinski. Synthetic aperture imaging using sources with finite aperture: Deconvolution of the spatial impulse response. *J. Acoust. Soc. Am.*, 114(1):225–234, July 2003.

[3] H. Desoky, M.Y. Abou-Bakr, and Y.M. Kadah. Reconstruction using optimal spatially variant kernel for B-mode ultrasonic imaging. In *Proceedings of the SPIE*, volume 5035, pages 147–153, 2003.

[4] J. Kormylo and J.M. Mendel. Maximum likelihood detection and estimation of Bernoulli–Gaussian processes. *IEEE Trans. Inf. Theory*, 28(3):482–488, May 1982.

[5] K.F. Kaaresen. Evaluation and applications of the iterated window maximization method for sparse deconvolution. *IEEE Trans. Signal Processing*, 46(3):609–624, March 1998.

[6] T. Olofsson. Computationally efficient sparse deconvolution of B-scan images. In *2005 IEEE Ultrasonics Symposium*, volume 1, pages 540–543, Sept. 2005.

[7] C.Y. Chi, J. Goutsias, and J. Mendel. A fast maximum-likelihood estimation and detection algorithm for bernoulli-gaussian processes. In *Proc. Int. Conf. Acoust., Speech, and Signal Process.*, volume 10, pages 1297–1300, 1985.

[8] F. Lingvall, B. Piwakowski, and G.M. Zang. *The DREAM (Discrete REpresentation Array Modeling) toolbox*. Freeware toolbox available at http://www.signal.uu.se/Toolbox/dream/.

[9] B. Piwakowski and K. Sbai. A new approach to calculate the field radiated from arbitrarily structured transducer arrays. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 46(2):422–440, March 1999.

[10] J.A. Seydel. *Ultrasonic Synthetic-aperture Focusing Techniques in NDT*. Research Techniques for Nondestructive Testing. Academic Press, 1982.

[11] V. Schmitz, S. Chalkov, and W. Muller. Experience with synthetic aperture focusing technique in the field. *Ultrasonics*, 38:731–738, 2000.

[12] F. Lingvall, T. Olofsson, and T. Stepinski. Synthetic aperture imaging using sources with finite aperture: Deconvolution of the spatial impulse response. *J. Acoust. Soc. Am.*, 114(1):225–234, July 2003.

[13] M. Soumekh. Reconessaince with ultra wideband uhf synthetic apreture radar. *IEEE Signal Proc. Mag.*, 12(4), July 1995.

[14] D.C. Munson Jr., J.D. O'Brien, and W.K. Jenkins. Tomographic formulation of spotlight-mode synthetic aperture radar. *J. Acoust. Soc. Am.*, August 1983.

[15] B. Bamler. A comparison of range-doppler and wavenumber domain sar focussing algorithms. *Transact. Geosci. Remote Sensing*, 30:706–713, July 1992.

[16] P.T. Gough and D.W. Hawkins. Imaging algorithms for a strip-map synthetic aperture sonar: Minimizing the effects of aperture errors and aperture undersampling. *IEEE Trans. Ocean Eng.*, 22:27–39, 1997.

[17] M. Soumekh. *Synthetic Aperture radar signal processing with* MATLAB *algorithms*. John Wiley & Sons, Inc, first edition, 1999.

[18] G.S. Kino. *Acoustic Waves: Devices, Imaging and Analog Signal Processing*, volume 6 of *Prentice-Hall Signal Processing Series*. Prentice-Hall, 1987.

[19] L. Schmerr. *Fundamentals of Ultrasonic Nondestructive Evaluation: A Modeling Approach*. Kluwer Academic / Plenum Publishers, 1998.

[20] A. Lhémery. Impulse-response method to predict echo-responses from targets of complex geometry. Part I: Theory. *J. Acoust. Soc. Am.*, pages 2799–2807, November 1991.

[21] P.T. Gough and D.W. Hawkins. Unified framework for modern synthetic aperture imaging algorithms. *Int. Journal of Imaging Systems and Technology.*, 8:343–358, 1997.

[22] F. Lingvall. *Time domain reconstruction methods for ultrasonic array imaging*. PhD thesis, Uppsala University, Signals and Systems, 2004.

[23] P. R. Stephanishen. Transient radiation from pistons in a rigid infinite planar baffle. *J. Acoust. Soc. Am.*, 49:1627–1638, 1971.

[24] H. Lasota and R. Salamon. Application of time-space impulse responses to calculations of acoustic field in imaging systems. *Acoustic Imaging*, pages 493–512, 1980.

Chapter 4

Three-Dimensional Ultrasound Imaging

by Tomas Olofsson

4.1 Introduction

Three-dimensional (3D) ultrasonic imaging has in the recent years become an important area of research. The main development has taken place in medicine, with a long tradition of 3D imaging in computerized tomography and magnetic resonance imaging, and 3D ultrasonic imaging is a natural next step for extending the possibilities for the clinician to monitor the human organs [1]. The ultimate goal is *real-time* 3D ultrasonic imaging that would help monitoring the movements and thus the physiological functionality of organs. The 3D technology is however still in the early stages of its development and few examples of clinical applications have been reported so far.

3D ultrasonic imaging also has a great potential in certain nondestructive testing (NDT) applications. One such application is the inspection of friction stirred welds (FSWs), in which a variety of flaws can be encountered, and 3D imaging would here be very useful for the inspection reliability. Unlike traditional 2D imaging, where cross-sections of the inspected volume are displayed and where the inspector must mentally reconstruct the involved volumetric structures, 3D images have the advantage of being seen from whatever angle the inspector may desire. 3D ultrasonic imaging is a natural continuation of our previous research focused on the high resolution imaging material flaws using ultrasound.

An important distinction between medical and NDT applications, such as FSW inspection, is that the latter typically do not involve moving structures and real-time requirements are therefore typically not as critical in NDT as they are in medicine; the speed at which the data can be acquired and presented is mainly a matter of avoiding bottle-necks in the production/inspection-line or a matter of making the inspection convenient for the operator. We may therefore expect that NDT imaging application will require less advanced equipment than those that are considered for 3D medical imaging and much of the equipment existing today can be used. However, an increase in computer demands compared to the systems used today is also to be expected.

3D ultrasonic imaging can generally be divided into three interrelated modular steps illustrated in Fig. 4.1: Data acquisition, 3D reconstruction, and 3D visualization. The data acquisition must ensure that sufficient information is gathered for faithfully reconstructing the inspected volume. Once the reconstruction has been performed, the user can choose an appropriate visualization mode to view the data.

Figure 4.1: The modular steps in a 3D visualization system

Several possible methods exist for each of the above mentioned steps and the choices should be based on design specifications that may involve system complexity, acquisition time, reconstruction time, image resolution etc. The choices of data acquisition and reconstruction methods are intimately related and in Sections 4.2, 4.3, and 4.4, acquisition methods are described along with their corresponding reconstruction methods. In the first of these sections, traditional line-by-line imaging techniques are introduced and later extended to 3D imaging. In Section 4.3 various methods based on the so-called synthetic aperture (SA) technique are described to-

gether with some accompanying methods for improving the signal-to-noise ratio (SNR). Section 4.4 briefly describes techniques related to tomography. In Section 4.5, the different possible visualization modes are described and, finally, in Section 4.6, a discussion on which methods of 3D ultrasonic imaging that can be suitable for the inspection of FSWs is given.

4.2 Data acquisition and reconstruction

3D ultrasonic imaging involves a geometrical conversion from the received data to a 3D coordinate system. The basic property that makes an ultrasonic transducer particularly suitable for 3D imaging is that it can measure the distance between the transducer and a scatterer through the time delay between transmission and echo reception. For this to be possible without positioning ambiguities, the data must be acquired from points that span a plane in front of the inspected volume, cf. triangulation.

The entity to be reconstructed is often considered to be the *reflectivity* as a function of position in space, $o(\mathbf{r})$.[1] The simplifying assumptions that are usually made are that objects in the region of interest can be synthesized by a set of point-like scatterers having different reflectivity (strengths) and that the mapping between the received data and $o(\mathbf{r})$ is linear. These approximations are valid under weak scattering, where the medium is assumed to be homogeneous with only small variations, but can be questioned otherwise. In particular, the linearity assumption is clearly violated whenever the reflectivity at some region is strong enough to completely shadow a region beneath. Then, regardless of the strength of any scatterer in this shadowed region, these scatterers' contributions to the received data will be zero.

For stationary objects, which is the typical case in NDT, the data does not need to be acquired simultaneously from all points on the plane and scanning can be used instead to obtain the necessary data. At one extreme, the data can be acquired using a single transducer that is mechanically scanned over the surface.[2] This, however, may require too much time to cover the volume of interest. At the other extreme, dense 2D ultrasonic arrays consisting of on the order of 10^4 elements can potentially be used. Since such arrays could provide us with a great amount of data at each reception, it would be possible to shorten the acquisition time significantly by using such arrays and, if flexibility and performance in terms of acquisitions speed, resolution etc are the primary design criteria, 2D arrays are generally the optimal choices in most applications.

We should note, however, that the conditions under which ultrasonic array systems operate are technically quite demanding. Sampling frequencies at 50 MHz and higher are not uncommon and, as mentioned above, the number of array elements can be quite large. Full waveform acquisition at all elements is therefore generally not feasible for large 2D arrays[3] because the involved circuitry (cabling and leads) then becomes difficult to manufacture and the array cumbersome to use.

A number of methods exist aiming at overcoming these technical difficulties and to obtain an acceptable compromise between performance and equipment cost. This compromise can consist in that only one or a few rows of the array are actively used at a time. Another alternative method is to use so-called *sparse arrays*. In such arrays, the distance between the elements is larger than half the wave length, $\lambda/2$, which is required to avoid so-called grating lobes, see below.

[1] Note however that in the majority of the publications in the research field, the entity of interest in the reconstruction is not clearly defined or mentioned at all.

[2] This is the mono-static case of synthetic aperture imaging described in Section 4.3.

[3] The largest fully sampled 2D array with full steering flexibility at transmission found in this literature study consists of 1024 elements [2].

The above mentioned technical compromises are discussed in the sequel. However, since much of the work on 3D ultrasonic imaging has gradually evolved from 2D imaging using 1D arrays, we here begin with a brief description of such arrays.

4.2.1 2D imaging using 1D arrays

The 1D arrays used for 2D imaging vary in shape and complexity. At transmission there are two distinctly different levels of flexibility of which the first is illustrated in Fig. 4.2 below:

- *Phased arrays* allow for focused and steered beams by the application of appropriate individual time delays, or a *focal law*, to the excitation impulses at transmission and at the received signals. These delays correspond to differences in sound propagation time between the elements and the focal point. The beam patterns can be further shaped by apodization (also known as tapering or shading), which means that the transmission from the nth element is weighted by a separate factor, a_n. The phased arrays are presently the ones most commonly used for imaging.
- *Array using arbitrary waveform excitation*, individual for each element. These allows for the most flexible wave field generation. They can simulate, as a simple special case, the operation of phased arrays and multiple focal zones can be created in one transmission by superimposing the excitations associated with separate focal laws.

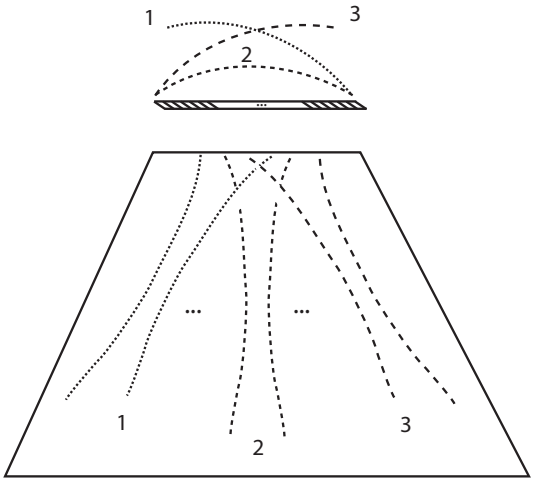


Figure 4.2: Illustration of line-by-line 2D image acquisition using phased arrays. Individual delays are applied to the elements to create simultaneous focusing and steering. The distance between the lines (1-3) and the array elements illustrate the delays applied to the excitation pulses at the array elements for the three steering angles.

The by far dominating strategy used when designing traditional ultrasonic imaging (2D) systems has been to create narrow beams so that the echo contributions from different scatter positions are as well resolved as possible already in the raw data format. The images are obtained line by line by steering the beam sequentially in different directions using phased arrays, along with focusing at some depth. The beam is steered both at transmission, as illustrated in Fig. 4.2, and at reception, using the delay and sum (DAS) technique as illustrated in Fig. 4.3. The focusing and steering at transmission and reception essentially mimic the behavior of an acoustical lens or a curved, geometrically focused aperture.

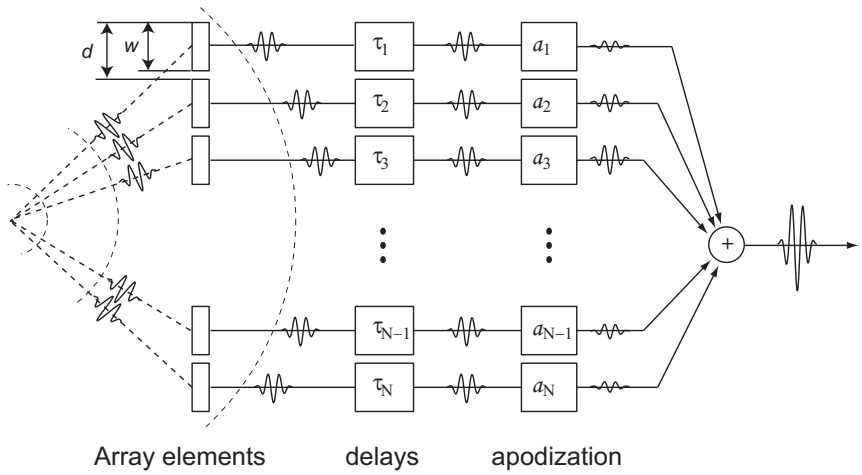


Figure 4.3: Illustration of receive focusing using the delay and sum (DAS) technique. The contributions from the focal point are summed coherently after given a separate apodization weight.

Beside the delays used in the focal laws, the critical parameters determining the shape of the beam pattern are the number of elements, N, the width of the elements, w, and the distance between the elements, d, and the apodization weights, $a_1, ..., a_N$, one for each element. These weights determine how strongly each element contributes to the transmitted signal. Or, at reception, how much gain is applied to the separate received signals.

For a linear phased array, these parameters uniquely determine the array's aperture function, $a(\xi)$ which describes the excitation strength as a function of position. This aperture function, in turn, determines the beam pattern. Note that we may have a different transmit aperture function, $a_T(\xi)$, from the receive aperture function, $a_R(\xi)$.

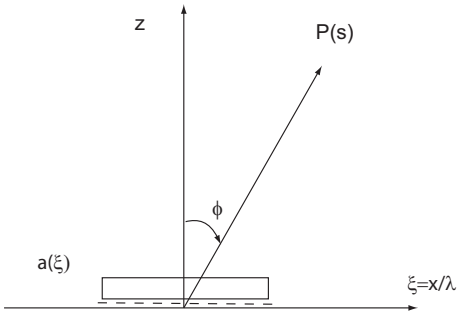


Figure 4.4: Coordinate system for relating the aperture function, $a(x/\lambda) = a(\xi)$ to the beam pattern, $P(s)$.

Consider the transmission of a continuous wave with excitation frequency, f_0, yielding a wave length $\lambda = c/f_0$, with c being the wave propagation speed. In the far field, the continuous wave

transmit beam pattern $P_T(s)$ is given by the Fourier transform

$$P_T(s) = \int_{-\infty}^{\infty} a_T(\xi) e^{i2\pi\xi s} d\xi \tag{4.1}$$

where $\xi = x/\lambda$ is the element location measured in wavelengths, $s = sin(\phi)$ with the angle ϕ defined in Fig. 4.4.

The receive beam pattern is similarly given by the Fourier transform of the receive aperture $a_R(\xi)$ and the round-trip, or pulse-echo, beam pattern is given by

$$P_{TR}(s) = P_T(s)P_R(s). \tag{4.2}$$

Below, some of the basic features of array beam patterns are illustrated. In the examples, the considered medium is water having a sound speed of 1500 m/s. The elements are rectangular with a length, L, that is assumed to be much larger than their widths, w, and we can then neglect any effect on the beam pattern caused by L.

Fig. 4.2.1 shows beam patterns for single element transducers of different widths. The aperture functions are "box functions" having sinc-functions as their corresponding Fourier transform. Elements with widths that are much smaller than the wave length produce almost spherical waves whereas larger elements create more narrow beams. Side lobes are clearly seen for the 1 mm element.

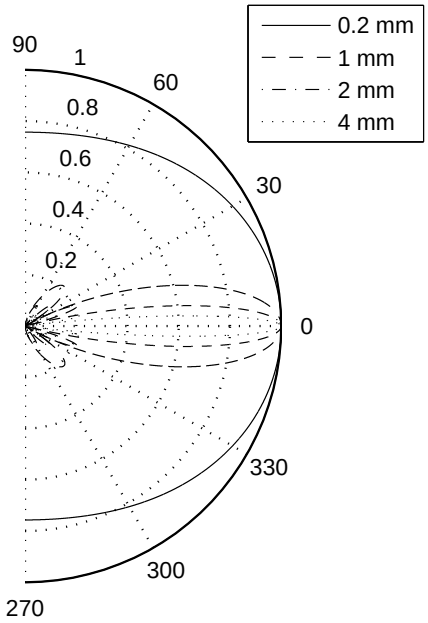


Figure 4.5: Beam patterns for single rectangular elements of different widths. The medium is water having an approximate sound speed of 1500 m/s and the frequency is 3 MHz which yields a wavelength $\lambda = 0.5$ mm.

The physical parameters w and d become fixed at the manufacturing and the user can only control the delays and the apodization weights. If a focal point has been chosen, the delays are given by purely geometrical considerations and, essentially, the only real choice becomes that of finding appropriate weights. By using uniform apodization, i.e. $a_1 = ... = a_N = 1$, the beams become maximally narrow, but this comes at the cost of relatively strong so-called *side lobes*,

next to the main lobe. Contributions from scatterers in the side lobe directions create artifacts in the images and it is therefore desirable to reduce these lobes.

The side lobes can be suppressed by using appropriate apodization weights. Most of the large number of apodization windows that have been proposed in the literature give extra weight to the central elements, resulting in an aperture that is effectively smaller than the full aperture.[4] Fig. 4.6, shows a comparison between the beam patterns of a 16 element array with uniform apodization and one with apodization weights given by a so-called Hanning window.

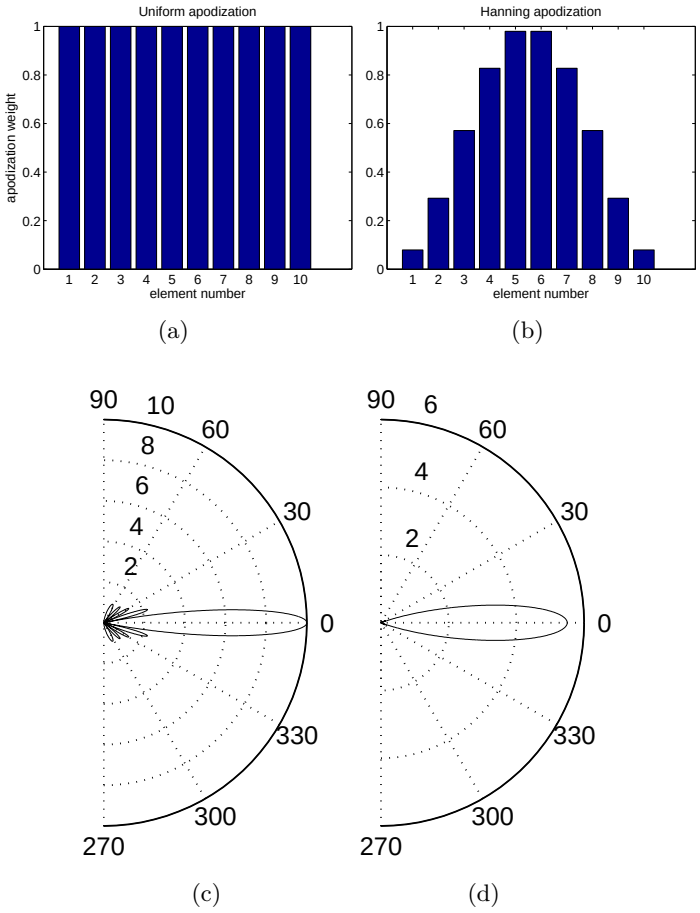


Figure 4.6: Illustration of the effect of apodization on beam patterns. (a) uniform weights (b) Hanning weights. (c) Beam pattern for uniform apodization. The main beam is maximally narrow but there are also strong side side lobes. (d) Beam pattern of the same array but with Hanning apodization. Now the main lobe is wider but the side lobes are much suppressed.

As seen in the figure, the suppression of the side lobe comes at the expense of a wider beam which results in worse resolution. Furthermore, apodization is typically implemented by dampers so the output power can only be decreased by applying it.

Other disturbances are the *grating lobes* that occur if the distance between the transducer elements, d is larger than half the wave length, $\lambda/2$, i.e., if the array is *under sampled*. This phenomenon is in close analogy with *frequency aliasing* that occur when a time signal is sampled at rate less than two times the maximum frequency found in the signal, the so-called Nyquist frequency.

[4]Of course, at one extreme only the central element is given any weight at all and the aperture than becomes identical to a single element aperture with beam patterns illustrated in Fig. 4.2.1

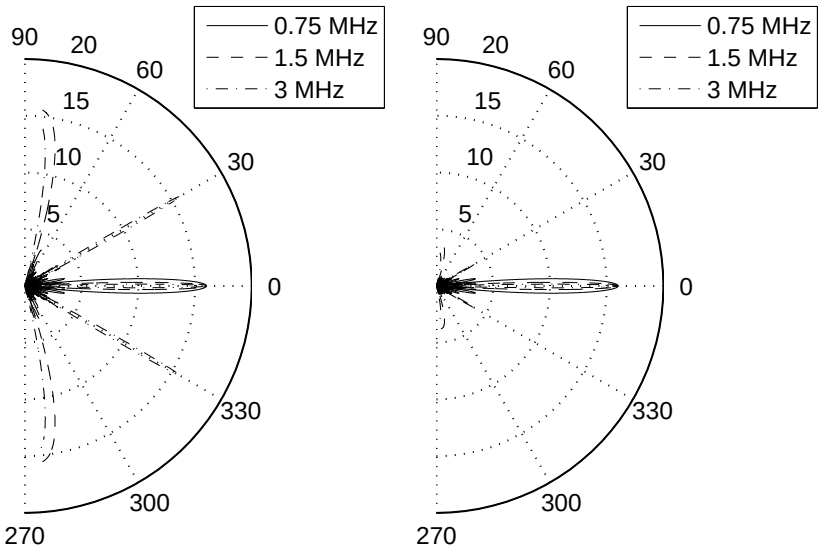


Figure 4.7: Beam patterns for arrays consisting of 16 rectangular elements. The medium is water having an approximate sound speed of 1500 m/s. (a) Beam patterns for array with very small width elements ($w = 0.1$ mm) separated by $d = 1$ mm. The patterns are shown for three frequencies: 0.75 MHz ($\lambda = 2$ mm), 1.5 MHz ($\lambda = 1$ mm), and 3 MHz ($\lambda = 0.5$ mm). For the two latter frequencies, the element separation is larger than $\lambda/2$ and as a result, grating lobes occur for these frequencies. (b) Same parameters as in (a) except that a wider element with $w = 0.8$ mm is used. The grating lobes are suppressed due to the directivity of the elements.

Grating lobes are often present in practical array system used for imaging because the problems of producing small enough array elements are considered to be greater than the benefit of avoiding these lobes.[5] Often the arrays are only slightly under-sampled, thus causing only one or at most two grating lobes at angles that are relatively far from the main lobe.

If the steering angle is small, such grating lobes will be suppressed as a result of finite size of the elements, w, which causes a directivity to the beam itself, see Fig. 4.7(b). The overall beam pattern, $P(\phi)$ for an array with equally spaced elements is the product of the beam pattern of a w-width single element, $P_w(\phi)$, and the pattern, $P_N(\phi)$, of an N element array with infinitesimally small elements.[6] In cases where the distance d is large, w can also be made almost equally large, thus yielding single element beam patterns that are fairly small at the angles at the angles where the grating lobes occur. For large steering angles, this does not necessarily hold. Note however that the angles of the grating lobes depend on frequency and grating lobes associated with wide band excitation are therefore "smeared" over an interval of angles in contrast to the main lobe which is in the same direction for all frequencies.

Except for one important difference, the focal laws used at transmission and reception are generally the same. The difference is that, at reception, as the echoes arrive back from the object, the delays can be modified in real time to ensure that the positions that instantaneously contribute to the signals are always in focus. This so-called *dynamic focusing* technique can be implemented in hardware and it is regularly used in ultrasonic arrays. Systems using dynamic

[5]This holds when the medium is water which is relevant if the inspection is to be performed in immersion. If contact testing is considered, the wavelengths of interest become larger because of the greater wave propagation speed in solids.

[6]This can be realized by noting that this aperture can be written as a convolution between a "box" of width w and a "spike train" consisting of N spikes. Since the overall beam pattern is the Fourier transform of the aperture function, the convolution translates to a product between the beam patterns.

focusing often transmit unfocused beams (steering only) and the focusing is performed in reception only. In this way the resolution becomes approximately uniform in depth. Dynamical focusing can only be obtained by electrical means or by using post processing provided that the signals from all elements have been stored; it cannot be achieved using geometric focusing.

The acquisition rate can be increased by reconstructing a few image lines at each reception. The transmit beams used in such systems are relatively wide and a set of delay-and-sum circuits are used to reconstruct several image lines in parallel using the *Explososcan* processing technique [3]. This processing technique is illustrated in Fig. 4.8 below.

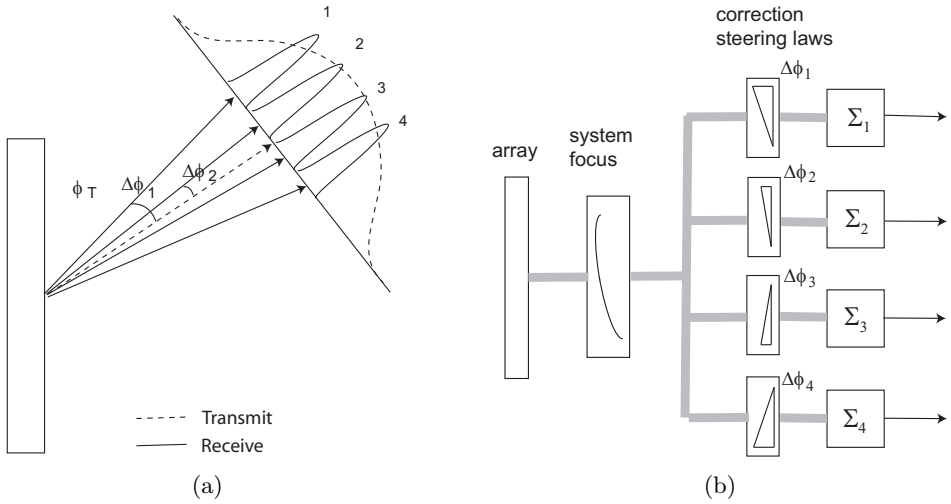


Figure 4.8: Illustration of the Explososcan parallel processing technique. (a) At transmit, a relatively wide beam is steered in direction ϕ_T. At receive, a number of slightly different focal laws corresponding to directions $\phi_T + \Delta\phi_1, ..., \phi_T + \Delta\phi_4$, are applied in parallel to create a set of image lines from one transmission. (b) In the implementation, a common focal law identical to the transmit focal law (system focus) is applied to all lines. Separate steering is subsequently applied, in parallel, to obtain the four different lines. The steering corresponds to the image lines' deviations from the transmit beam direction. These steering laws can be implemented with a simpler technique (tapped delay lines) than the system focus.

4.2.2 3D imaging using mechanically scanned 1D arrays

In practice, provided that mechanical scanning is acceptable for the application, the dominating method to acquire 3D data is to use a combination of mechanical scanning and electrical scanning using 1D arrays of the same type that are used for ordinary 2D imaging. The data volume is obtained slice by slice, with the slices obtained as described in Section 4.2.1. The obvious advantage of this approach to collect the data is that it minimizes the need for advanced post processing.

In medicine, the arrays are often hand-held and the scanning is performed manually. The 3D reconstruction requires that precise positioning information accompanies the received data and for hand-held devices the solutions involve either triangulation using acoustical devises or magnetic sensors [1]. In NDT applications, it is more common to use mechanical scanning and the transducer's position is usually accessible from the scanner quite easily.

The 3D acquisition is illustrated in Fig. 4.9. One image plane is acquired at each scanning position. To obtain well separated planes, i.e. to obtain a good *elevation focus*, the elements in the array are geometrically focused at some range, f. At distances closer than f, the wave have not converged and further away the waves are diverging. Thus, outside a certain range

interval, the elevation resolution is usually poor. This problem can be addressed using more flexible arrays as described in the following section.

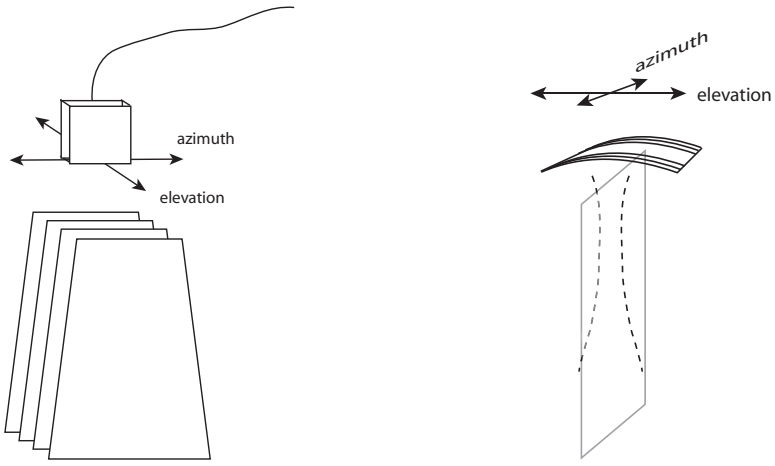


Figure 4.9: Traditional scanning of a volume plane by plane using a 1D array. (a) The array is mechanically scanned acquiring one plane at each position. The image from one plane is obtained by electronically steering a beam in a number of angles. (b) The separation between the planes is accomplished by using elements that are geometrically focused in elevation.

1.25D-1.75D arrays

The 1D arrays consist of a set of narrow strips with widths w and length L. The strips are curved to create the geometrical focus. More flexible arrays are required to for addressing the elevation focus issue mentioned above. This can be done by cutting the strips to get a number of rows of elements and connecting the elements in these rows to separate cables. The connections can be done in a number of ways. Following a nomenclature in Wild et al, [4] and that has been adapted by several authors, the commonly used array structures are referred to as 1D, 1.25D, 1.5D, 1.75D, and 2D arrays and they are defined as follows:

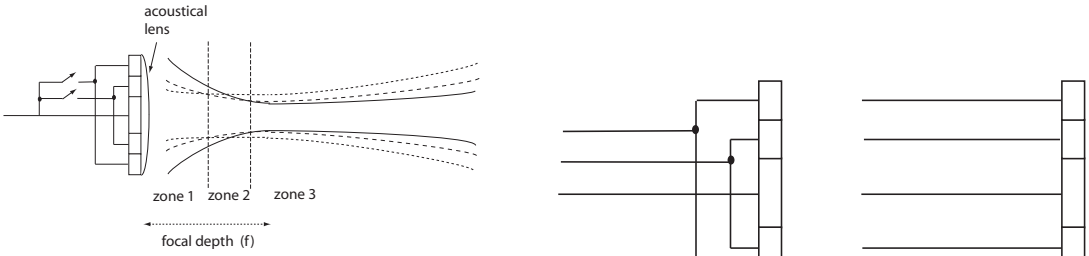
- 1D: Elevation aperture is fixed, and geometrically focused at a fixed range (as described earlier).
- 1.25D: Elevation aperture is variable, but focusing remains static.
- 1.5D: Elevation aperture, apodization, and focusing are dynamically variable, but symmetric about the centerline of the array.
- 1.75D: A 1.5D array without the symmetry constraint. Elements are large (several wavelengths) in elevation, so very little steering is possible.
- 2D: Elevation geometry and performance are comparable to azimuth, with full electronic apodization, focusing, and steering.

A common theme for the 1.XD arrays is that the elements are fewer and larger in the scanning direction than in azimuth. In this way the total number of elements may become feasible and

some of the practical manufacturing and handling problems of 2D arrays mentioned earlier can be avoided.

The use of 1.25D arrays is one of the simplest methods to extend the range interval in which the elevation resolution is satisfactory. The idea is based on the following observation: Suppose that two strip elements, A and B, have the same focal range, f, but different lengths (apertures), L_A and L_B, with $L_A < L_B$. Then B has a better elevation resolution than A at ranges beyond a point just before f, whereas the opposite holds at shorter ranges. The measurements from both A and B can be blended to extend the range interval with good resolution.

The idea is illustrated in Fig. 4.10(a) for a 1.25D array consisting of five rows. Activating all segments gives the narrowest beam at the focal point and further away. Closer to the array, the smaller apertures that are obtained by disconnecting one or both of the switches give more narrow beams than using the full aperture. Note that due to the electrical disturbances that occur when switching between the different states, the measurements require as many transmit-receive cycles per beam as there are states (three in this example).



(a) 1.25D array (b) 1.5D array (c) 1.75D array

Figure 4.10: Illustration of 1.25D, 1.5D, and 1.75D arrays. (a) The elements of the 1.25D array use no electrical focusing but improve the resolution in the near field by reducing the aperture size, something that is controlled by the switches. The smallest aperture yields the narrowest beam in zone 1, the medium aperture in zone 2 and the full aperture yields the narrowest beam in zone 3. (b) Individually delayed excitation can be applied to element pairs in a 1.5D array. The focus is restricted to lie along the symmetry axis whereas in (c) the 1.75D array, separate access is given to all elements thus allowing also for steering.

In Fig 4.10(b) a 1.5D array is illustrated. It allows for the separate excitation of the element pairs that are found symmetrically around the central element. By applying time delays to these, a focal law can be used to control the elevation focus. Dynamical focusing can be used to accompany the dynamical focusing used in azimuth.

The extension from 1.5D to 1.75D arrays is obtained by removing the connection between the symmetrical pairs which allows for applying separate delays to all elements and, thus, steer the beam. The flexibility of steering in the elevation direction is still somewhat limited due to the relatively long strips used.

4.2.3 2D arrays with restricted access to the elements and sparse arrays

The arrays mentioned above are all less flexible in elevation focusing compared to azimuth and typically have worse resolution in this direction. To obtain equal focusing properties and enabling electrical scanning in both directions, 2D arrays are required. These are manufactured with elements of the same size and number in both directions but the complexity is usually reduced by using switches that activate only a small subset of the available array elements

during transmission and/or reception. In this way the need for massive data transfer using a large number of cables is reduced.

The use of different apertures at transmission and reception provides flexible means to shape the beams. The received beam can be designed to compensate for shortcomings of the transmit beam. For instance, if the transmit beam has strong grating lobes in a certain direction, the receive beam can be designed to canceling the lobe by placing a null in this direction. This technique was used by von Ramm et al. [5] for 1D arrays and later adopted to 2D arrays by Smith et al. [6, 7].

There is an endless number of transmit- and receive element configurations that can be used and search for the optimal choice is a open research field. One relatively simple configuration that illustrates the concept is the so-called Mills cross, see Fig. 4.11. It was used in [6, 7] in which one of the first developed real time 3D ultrasonic imaging prototypes is described.

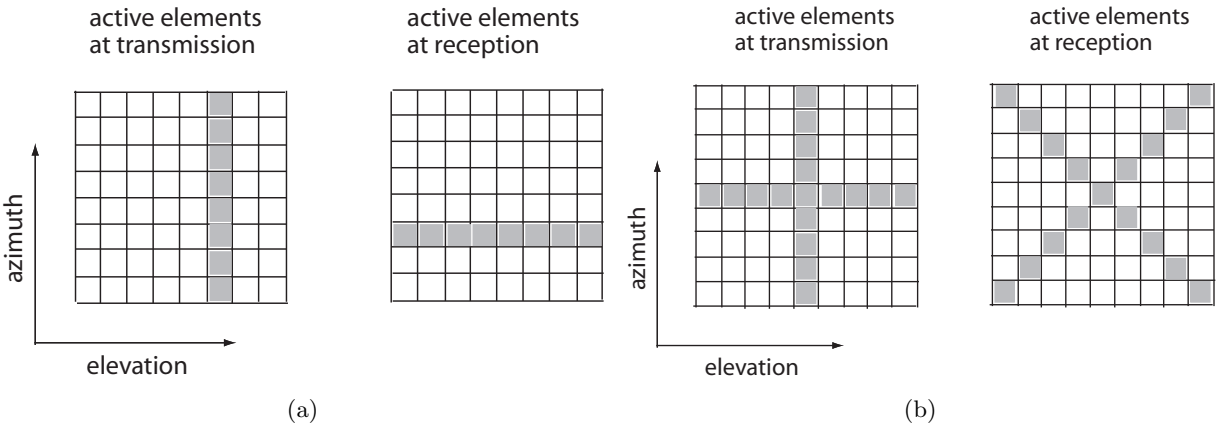


Figure 4.11: Two versions of the Mills cross technique. (a) At transmission, one column of the array is activated. This results in a beam that is narrow in azimuth but wide in elevation. At reception, a row is used for focusing and steering in elevation. The combined result is a beam that is narrow in both elevation and azimuth. (b) A version of the Mills cross that further improves the beam patterns. This version was used in [6].

An important tool when designing combinations of transmit and receive beams using reduced set of elements is the concept of *effective aperture* [8]. The combination of transmission and reception gives rise to an overall round-trip beam pattern. The effective aperture is a hypothetical receive aperture that would yield this round-trip beam pattern if the transmit aperture was a point source [9].

Recall $P_{TR}(s)$ in eq. (4.2) in Section 4.2.1. We can define the effective aperture, $a_E(\xi)$, as the aperture that has $P_{TR}(s)$ as its Fourier transform. Since products in the Fourier domain translate to convolutions in the spatial domain, we have that the effective aperture is given by the spatial convolution between the transmit- and the receive apertures.

$$a_E(\xi) = a_T(\xi) * a_R(\xi). \tag{4.3}$$

The concept of effective apertures is readily extended to 2D arrays and is useful in the design of beamformers.

The presence of grating lobes described earlier is one of the problems that can be treated by a suitable combination of transmit and receive apertures. We saw in Fig. 4.7 that strong contributions could be seen at one or a few angles away from the intended steering angle and the position of these lobes are determined by the separation between the elements. Between these

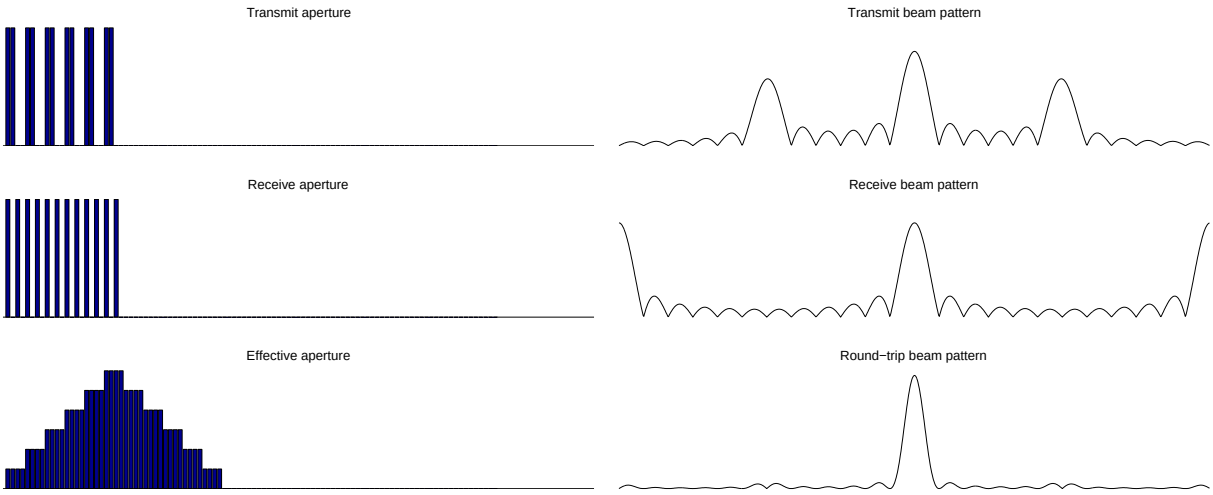
maxima, there were also a number of minima (zeros). In [5], von Ramm et al. suggested the use of different spacing of elements at transmission and reception, making it possible to place the zeros of the receive beam pattern to destructively interfere with the grating lobes in the transmit pattern, and thus obtaining a round-trip beam pattern that are almost free from grating lobes.

Sparse arrays

The condition for avoiding grating lobes in an array is that the inter-element spacing, d, is less than $\lambda/2$. Arrays of a certain size that have fewer elements than required by this condition are said to be "sparse". Some arrays are manufactured to be sparse. However, effectively sparse arrays can also be simulated using dense arrays, if they actively use only subsets of the elements in which the spacing between active element do not fulfill the condition.

The design of sparse arrays follows the same line as described above: An effective aperture corresponding to a desired round-trip beam pattern is calculated. The goal is then to fit a combination of transmit and receive apertures to this effective aperture under some constraints that may involve minimum spacing between the elements or maximum number of elements. The optimization problems are difficult and there are rarely any simple closed form solutions to these problems. However, some practically useful design strategies have been proposed:

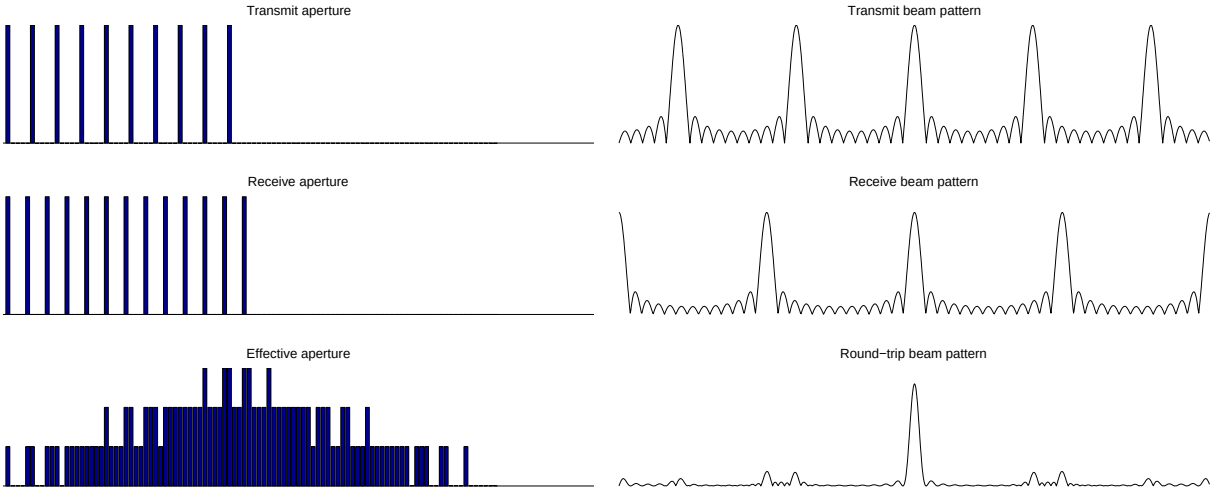
- Randomly placed elements [10]. A main theme for all strategies is to destroy the periodic patterns in the array layout to avoid strong grating lobes and a random layout is of course unlikely to create any significant periodic patterns.
- Rectangular interpolation [11]. One of the apertures is given a periodic sparse pattern with one element followed by gaps of $q-1$ element. The other aperture periodically repeat groups of q elements followed by q missing, see an example in Fig. 4.12.
- Vernier interpolation [11]. The array can be designed by analogy with the linear vernier scale. If one aperture has spacing pd, with p being an integer, and the other has a spacing of $(p-1)d$, then the spacing of the effective aperture becomes d. The integer p determines the sparsity of the array. One example with $p = 5$ is given in Fig. 4.13
- Synthetic aperture [11]. Here the transmit elements are individually excited and the backscattered energy is recorded and stored for each receive element. A beam can be synthesized off-line by appropriately delayed and weighted signals from every transmitter-receiver combination. Synthetic aperture techniques are discussed more in Section 4.3.



(a) Transmit and receive aperture using rectangular interpolation

(b) Transmit-, receive-, and round-trip beam patterns.

Figure 4.12: Illustration of the rectangular interpolation design strategy for a 1D array.



(a) Transmit and receive aperture using Vernier interpolation

(b) Transmit-, receive-, and round-trip beam patterns.

Figure 4.13: Illustration of the Vernier interpolation design strategy for a 1D array.

The beam patterns can be further shaped by using appropriate apodization to reduce the size of remaining side- and grating lobes.

For 2D arrays, the apertures depend on both x and y and have far more degrees of freedom than the 1D apertures. Thus, the optimization of 2D transmit-, receive aperture is computationally more demanding than for 1D arrays. However, the problem can be greatly simplified by restricting the apertures to be separable, i.e., requiring that

$$a_T(x,y) = a_T(x)a_T(y) \qquad \text{and} \qquad a_R(x,y) = a_R(x)a_R(y). \tag{4.4}$$

As a consequence, the effective aperture $a_E(x,y)$ also becomes separable

$$a_E(x,y) = a_E(x)a_E(y), \tag{4.5}$$

and the problem can then be reduced to that of optimizing 1D apertures using for instance the strategies described above. The separable apertures tend to produce beam patterns containing "squarish" artifacts. These can be suppressed to some extent by applying radially symmetric apodization weights.

4.3 Diffuse excitation and synthetic aperture imaging

One drawback with the traditional line-by-line approach is the relatively long time that is often required for covering the volume of interest. This is because the data acquisition rate is limited by wave propagation velocity in the object. We saw earlier in Section 4.2.1 that several image lines can be obtained from a single transmission (Explososcan). In synthetic aperture (SA) imaging this idea is elaborated further. In SA, transmissions that excite larger parts of the volume, so-called *diffuse excitation*, are used and the reconstruction consists in applying separate focal laws for each point in the object. Delay line implementations, as in Explososcan, cannot be used for the reconstructions of regions insonified by very wide beam since these implementations rely on approximations that hold only in a relatively small angle interval around the transmit beam (around 0.4 rad). Instead, SA requires that signals are sampled from the elements and and stored separately and the receive beamforming is performed off-line.

SA imaging was invented in radar more than 50 years ago. There, a small wide beam antenna, typically mounted on an aircraft, transmitted and received at positions along a line and the contributions from each scattering position were coherently summed. The SA technique was adopted to ultrasonic applications in the mid-seventies [12–14] and is still today an active research area. Radar most often involve a single scanned antenna, the so-called *mono-static* case, whereas arrays are routinely used in ultrasound (multi-static case). As a consequence, a number of different techniques have evolved in ultrasonics. They can be categorized as follows:

- Synthetic receive aperture (SRA)[15, 16]. This rather tedious method was developed for low complexity hardware that was designed to acquire data from only one receive element at a time. If N_{rx} denotes the number of receive elements, one image line is created in N_{rx} transmissions/receptions. At each transmission, the beam is steered in the desired image line direction and the data is recorded at one receive element each time. From the recorded data, the scatterer positions along the line are restored by coherently summing their contributions.

- Synthetic transmit aperture (STA)[17]. STA can be used with arrays that acquire simultaneously all receive channels. One entire image plane is reconstructed from N_{tx} transmissions, where one transmit element is used each time. Low resolution images are reconstructed from each transmission and these are summed after the N_{tx} transmissions to form a final high resolution image.

- Synthetic transmit and receive apertures (STA/SRA). In its basic form one transmission per array element is performed and reception is performed using a number of elements, but not all.

Most modern arrays systems allow for simultaneous acquisition from several elements. Since the reconstruction quality always will benefit from more data, there are no reasons to use SRA unless the hardware have such restriction. The same comment holds also for the SRA-part in the combination STA/SRA; if the system allows for the reception from all element, then this should be used otherwise information will be thrown away. Below we therefore only consider STA but before we describe this technique, let us consider the mono-static case to understand

some of the basic properties of general SA. The presentation below which describes SA for 2D imaging follows relatively closely that by Nikolov [18].

Consider an array with elements on the x-axis. A small element at position x_i transmits almost cylindrical waves and a point scatterer at coordinates (x_p, z_p) reflects the wave back to the same element.

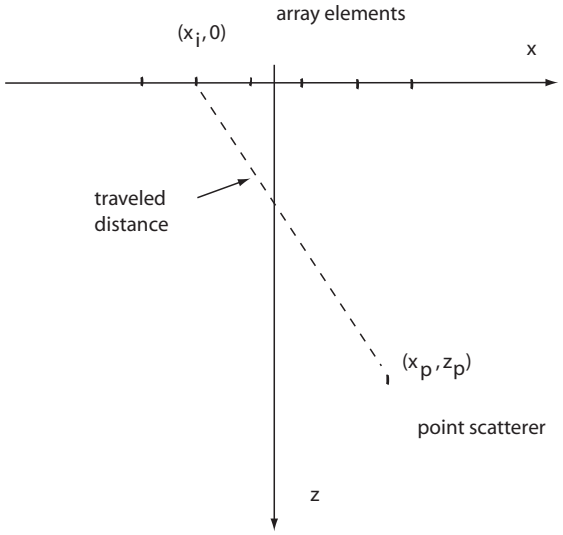


Figure 4.14: The coordinate system used in the presentation.

We transmit at time $t = 0$. The round-trip delay between transmission and reception of the back-scattered wave is

$$t_p(x_i) = \frac{2}{c}\sqrt{z_p^2 + (x_i - x_p)^2}. \tag{4.6}$$

If the electrical impulse response of the element is $h(t)$ and the scatterer is in the far field and has strength o_p, then the received signal is

$$y(t, x_i) = o_p h\left(t - 2\frac{\sqrt{z_p^2 + (x_i - x_p)^2}}{c}\right). \tag{4.7}$$

For a collection of scatterers in the medium, we have the superposition of contributions

$$y(t, x_i) = \sum_p o_p h\left(t - 2\frac{\sqrt{z_p^2 + (x_i - x_p)^2}}{c}\right). \tag{4.8}$$

The reconstruction in SA is basically an off-line imitation of the DAS concept used in phased arrays. The differences are that the data has been recorded and stored in memory and, for the mono-static SA case, that the appropriate time delays to obtain coherent summation is twice that used in the phased array. The reason for this additional factor two is that we now consider the round-trip delays and not only the delays associated with the forward- or backward sound path.[7]

The reconstruction of the object function at (x_p, z_p) is performed as

$$\hat{o}(x_p, z_p) = \sum_i \tilde{y}(t_p(x_i), x_i), \tag{4.9}$$

[7]This factor two is the one appearing in eqs. (4.6)-(4.8).

where $t_p(x_i)$ is the round-trip delay in eq. (4.6) and $\tilde{y}(t, x_i)$ is a pre-processed version of the received signal $y(t, x_i)$. The pre-processing may involve a deconvolution filter to remove the smearing effects of the electrical impulse response or, more commonly, a filter matched to the transducer impulse response:

$$\tilde{y}(t_p(x_i), x_i) = y(t_p(x_i), x_i) * h(-t). \tag{4.10}$$

The most important feature of the matched filter processing, which can be seen as robust version of deconvolution, is that it restores the phase of the signals so that the peak occurs correctly at time $t_p(x_i)$.

Since $y(t, x_i)$ is a discretely sampled signal and since eq. (4.9) prescribes summation of a continuous function, interpolation is required to implement (4.9) correctly.

Several implementations of the reconstruction exist. For instance, it can be performed in the spatial Fourier domain [19] with the advantages of faster computations and that the interpolation mentioned above becomes a natural part of the algorithm.

If the single element is scanned over a large synthetic aperture it can been shown that the later resolution is depth independent and is determined only by the width, w, of the element:

$$\delta_x \leq \frac{w}{2}. \tag{4.11}$$

If the aperture size is taken into account, something which is more realistic, the mono-static SA produces images of resolution that is not identical but not far from the images obtained from phased arrays. This should be expected since the technique essentially tries to imitate phased array processing. Differences occur in that the resolution of SA is somewhat better but with slightly larger side lobe levels. However, by introducing suitable apodization weights in eq. (4.9), these differences can be made rather small. The main difference remaining is that the condition for avoiding grating lobes in mono-static SA is that $d < \lambda/4$ instead of the usual $d < \lambda/2$ condition that apply for phased arrays.

Even in its basic mono-static form, SA competes favorably in terms of acquisition speed with phased arrays covering the same aperture. Mono-static SA requires N_{tx} transmissions to form an image and phased arrays require N_l lines to create the same image. An approximate formula [20] for this number that is determined by the lateral resolution of the system is $N_l \geq 1.5N_{tx}$. Thus, SA completes one image frame in about 67% of the time used by an equivalent phased array.

4.3.1 Synthetic transmit aperture (STA)

In STA, one element transmits and the reception of the backscattered wave is performed using all elements. The process is repeated, with a new transmit element for each acquisition until all elements have transmitted, resulting in overall N_{tx} transmissions. The full waveforms are sampled from the receiving elements and stored.

Because of the small size of the transmitting elements, the entire volume is insonified at each transmission. Thus, scattering information from the entire volume is provided at a single transmission and, in principle, the received data can be used to reconstruct this entire volume. This results in N_{tx} reconstructed images of comparatively low resolution. To form the final high-resolution image, these low-resolution images are added.

DAS is again used for the reconstruction of the low resolution images but with different time delays than in eq. (4.6). If the array's center is at $(0, 0)$, the beamforming applied to reconstruct

(x_p, z_p) is given by

$$\tilde{o}_i^p = \sum_{j=1}^{N_{rc}} a_{ij} \tilde{y}_{ij}(t - \tau_{ij}^p) \tag{4.12}$$

where $\tilde{y}_{ij}(t)$ is the received signal[8] at element at x_j resulting from a transmission at an element at x_i, and a_{ij} is the apodization. The delays are given by [18]

$$\tau_{ij}^p(t) = \frac{1}{c}\sqrt{(x_p - x_i)^2 + z_p^2} + \frac{1}{c}\sqrt{(x_p - x_j)^2 + z_p^2} - \frac{2}{c}\sqrt{x_p^2 + z_p^2}. \tag{4.13}$$

The high-resolution image is obtained by a summation

$$\hat{o}^p = \sum_{i=1}^{N_{tx}} \tilde{o}_i^p. \tag{4.14}$$

Nikolov and Jensen have described how to produce a fast stream of high resolution images while still only using the above mentioned reduced set of transmit elements [18]. The technique recursively uses data obtained for the N_{tx} latest transmitted elements. The elements are transmitted in a cyclic scheme that repeats the transmission from a certain element every N_{tx}th transmission. The method is useful when the inspected objects involve moving structures since it gives the possibility to track fast movements in large regions of interest but it is of limited interest when the object is stationary.

The images obtained using STA have about the same resolution as those obtained by phased arrays having the same aperture. Furthermore, STA opens up for the possibility to significantly reduce the acquisition time, which is directly proportional to the number of firings. Full size images can be created even if only a subset of the available elements are used. By appropriate design of the receive aperture we can maintain the same effective aperture as an equivalent phased array and, thus, maintain the same resolution.

Lockwood and Foster [21] have described an STA setup that uses only five transmissions to form high resolution images at very high rates. An analysis of the method showed that the price paid for reducing the acquisition time is poor SNR. The STA processing involves accumulation of several measurements whereas phased arrays require only one measurement per reconstructed line. Assuming that all N_{tx} transmit elements are used in STA and that the same noise level is present in both STA and phased array measurements, STA will accumulate up to a noise level that is $\sqrt{N_{tx}}$ times the one found in phased array measurement. The situation becomes even worse if fewer than N_{tx} elements are used (sparse STA), see [21] for details.

4.3.2 Increasing the SNR in systems that are using diffuse excitation

As seen above, one drawback with diffuse excitation is related to the weak signals compared to focused beams. To improve the SNR of the signals, the transmitted energy must be increased. The obvious way to do this is to increase the amplitude of the signal. In medicine, a large increase of the output levels have been prevented by concerns about possible tissue damage. In NDT application, this should be a minor concern. However, ultrasound systems always have an upper limit on how much power that can be delivered by the elements, a limit that may depend on nonlinearities in the transducer or properties of the electrical system supporting the array.

Even when such an upper amplitude limit has been reached, a further increase in the transmitted energy can still be done either by distributing this energy *spatially* or *temporally* or

[8]The signals may be pre-processed as mentioned earlier for the mono-static SA.

combinations thereof. Spatial distribution of the energy can be achieved by simulating one single strong element using several elements. This simulated element is called a *virtual source*. Temporal distribution can be achieved using so-called *pulse compression* techniques. These two techniques are described in more detail below.

Virtual sources

Virtual sources [21–23] can be created using either a geometrical focused transducer or an electronically focused array at transmission. At the focal point, the waves travel through approximately the same point and, thus, the focal point can be regarded as a strong point source.

One advantage of virtual sources over the pulse compression techniques described below, is that it does not rely on the use of arbitrary excitation waveforms meaning that ordinary phased arrays can be used and the electronics used at transmission can be kept simple. A disadvantage is that the model is over-simplified. The wave-field propagating from the focal point is not identical to the field from a point source and the conditions required in SA are violated, see [24].

Spatial encoding for STA measurements

Ciao et al. [25] have described a method for increasing the transmitted power in STA measurements by firing several elements at the same time. Consider a sparse STA using a set of M transmit elements. Instead of sequentially transmitting one of these elements at a time, as in ordinary STA, M different transmissions are used and in all these transmission all members in the set M are active. Different weights are given the transmissions each of the M times. Altogether there are M^2 weights that can be organized into a matrix $\mathbf{Q}_M$ as

$$\mathbf{Q}_M = \begin{pmatrix} q_{11} & \cdots & q_{1M} \\ \vdots & \ddots & \vdots \\ q_{M1} & \cdots & q_{MM} \end{pmatrix}, \tag{4.15}$$

where q_{ij} is the weight applied to element j at transmission i. These weights must be chosen so that $\mathbf{Q}_M$ becomes invertible.

Let $r_{ij}(t)$ denote the received signals at element j at the ith transmission. Furthermore, let us form the column vector $\mathbf{r}_j(t) = (r_{1j}, ..., r_{Mj}(t))^T$, of samples received by element j. Since these signals are nothing but weighted sums of the signals $y_{ij}(t)$ that were used in the beamformer in STA, see eq. (4.12), we can write $\mathbf{r}_j(t)$ as (encoding)

$$\mathbf{r}_j(t) = \mathbf{Q}_M \mathbf{y}_j(t) \tag{4.16}$$

where $\mathbf{y}_j(t) = (y_{1j}, ..., y_{Mj}(t))^T$, i.e., a column vector consisting of the samples received at time t by receiver j during M single element firings. Since $\mathbf{Q}_M$ is invertible we can decode $\mathbf{y}_j(t)$ as

$$\mathbf{y}_j(t) = \mathbf{Q}_M^{-1} \mathbf{r}_j(t) \tag{4.17}$$

and from these signals we can proceed with the STA beamforming as described earlier.

By choosing $\mathbf{Q}_M$ to be a so-called *Hadamard* matrix, the decoding becomes particularly simple. The element in Hadamard matrices are either $+1$ or -1 and the inverse is simply a scaled version of the transposed matrix: $\mathbf{Q}_M^{-1} = \frac{1}{M}\mathbf{Q}_M^T$. Both encoding and decoding thus involve only phase shifts (and sums in the decoding step), something that simplifies the implementations.

The above described spatial encoding gives an SNR improvement of $M^{1/2}$ compared to ordinary STA. In summary, the spatial encoding can give us back the loss in SNR caused by the accumulation of noise over several measurements in STA compared the phased array measurements and this gives the user better means to trade between acquisition time and SNR.

Coded excitation and pulse compression

A third way of improving the SNR is to distribute the energy over time, instead of over space as for virtual sources and spatial encoding. This is done by exciting the elements with long input signals. This idea is described below.

The pulses transmitted in an ultrasonic imaging system can often be fairly well approximated as a sine wave of frequency, f_0, that is weighted with an envelope function (time-window) of duration T yielding a bandwidth of approximately $B \approx 1/T$. Since a high bandwidth, B, is desirable for good axial resolution, T is usually made quite short. This is usually achieved by damping the oscillations of the piezo-electric crystal and it results in a transmitted wave that carries little energy.

If the system allows for arbitrary waveform excitation, the duration T can be extended while still maintaining the bandwidth. One simple method of doing this is by using frequency swept signals; the bandwidth of such a signal is determined by the start- and stop frequencies and the duration is determined by the rate of the sweep. Since this rate can be chosen arbitrarily small, the duration T can be made arbitrarily long. Thus, it is possible to transmit long duration signals that have wide frequency bands.

Pulses can be described using their so-called time-bandwidth product, TB, which for the first mentioned pulse is close to one and for the frequency swept signal may be much larger than one. Signals sharing the property that their time-bandwidth product are significantly larger than one are called *pulse compression signals*. Compared to a system using a certain signal with $TB \approx 1$, a system using a signal with the same B but with a larger TB can potentially gain in SNR with a factor TB [26].

The use of long duration pulses introduces a problem of poor axial resolution. This problem is solved by recovering (compressing) the received signals. This compression is made by filters that are matched to the excitation signals. See appendix 4.6 for details on pulse compression and how the matched filtering is performed.

4.4 Tomographic methods

Tomographic methods seek to reconstruct the interior of objects by illuminating these from all angles and measuring the projection of the objects in different planes. In classical computerized tomography using X-rays, it can be assumed that the involved beams are very narrow and that they travel along straight lines. A fundamental problem of using ultrasound for tomography is that the sound is refracted and diffracted to a much higher extent than X-rays. *Diffraction tomography* (DT) is a version of tomography that takes the whole wave fronts into account in the reconstruction and therefore is better suited for ultrasound measurements. DT come in several versions of which *reflection* DT is perhaps the most relevant for most applications.

In its basic form, DT uses planar harmonic waves at transmission and the measurements are performed at a series of positions enclosing the object of interest.[9] This particular setup allows for rather simple Fourier domain reconstruction methods but is unfortunately in many application rather impractical for a number of reasons: First, the use of only a single frequency at the reconstruction leads to comparatively poor resolution images and wide-band excitation has been proposed to solve this problem [27, 28]. Second, the requirement of plane wave excitation is difficult to realize with transducers of finite size and this requirement has been relaxed so that DT can be adopted to more practical array setups [29]. Third, the geometry of the objects may

[9] For 2D objects, the enclosing is a closed curve and for 3D objects, the enclosing is a closed surface.

not allow for the access from all angles and for this reason DT for limited angle coverage has been proposed [30, 31].

We should note that DT and DAS based techniques have historically evolved much in parallel. The theoretical framework of DT has generally been expressed in the Fourier domain and, although the rather theoretically advanced material has been well understood by physicists, it has perhaps not been as easily grasped by practically oriented researchers. Furthermore, some of the underlying assumptions of DT, such as those mentioned in the above paragraph, may have been too restrictive for a practitioner to accept. It is interesting to note that as DT has developed to be more practically oriented, the differences between DT and DAS based methods have become small. See for instance [32] and [28] where the authors point out how the Fourier based techniques can be well approximated using DAS when broadband signals are considered.

4.4.1 Limited diffraction beams

The last few years there has been an increased interest in so-called limited diffraction beams (LTBs) [33]. Methods that use LTBs and that resemble diffraction tomography have been developed.

LDBs are beams for which the wave fronts are constant as the wave propagates through the medium, a property that makes them suitable for use in imaging systems since the spatial invariance allows for using Fourier techniques in the reconstruction. Theoretically, limited diffraction beams can only be produced with an infinite aperture but they can be approximated using a finite aperture over a large depth of field.

For harmonic excitation, i.e., single frequency excitation, LDBs can be produced relatively simply using annular arrays. In this case the beams are created by applying a set of suitable apodization weights. Note that annular arrays require mechanical 2D scanning and the data acquisition therefore in its basic form is quite inefficient for 3D imaging. However, because of the flexibility of 2D arrays to simulate the behavior of essentially any kind of transducer, LDBs can also be produced by such arrays.

High resolution imaging systems generally require the use of broadband excitation. The apodization weights for creating the above mentioned harmonic beams are frequency dependent and, as a consequence, broadband LDBs require that the elements can be excited using separate waveforms. See for instance [34] for an example of individual excitation signals for a 10 ring elements in an annular array to create so-called X-waves.[10] LDBs cannot be created using 1D arrays but it has recently been shown how good approximations of such beams can be produced using 1.75D arrays with 10 rows or more [35].

An interesting application of limited diffraction beams is found in [36] and [37] where high frame rate 2D and 3D imaging using LDBs is described. In the proposed technique, a planar broadband wave pulse is transmitted and signals are acquired at all receiving elements on a 2D array. The LDBs are formed off-line using the stored data, very much similar to SA techniques, and images can be reconstructed using only one transmission. One difference between the technique using LDBs and SA is that SA is essentially based on excitation of small elements yielding diverging beams which results in poor SNR whereas the use of planar waves in the former gives a stronger excitation and a higher SNR is thus possible.

[10] The name comes from the X-shape of the propagating wave front.

4.5 Visual presentation

Provided that the acquisition and the reconstruction of $o(\mathbf{r})$ have been performed properly, the last step in 3D ultrasonic imaging is the visual presentation. There is no single obvious way of displaying 3D data on a 2D screen and different methods have been proposed. The three dominating methods for presenting 3D data volumes are (i) *multi-planar rendering* (ii) *surface based rendering* (iii) *volume rendering* [38]. These methods are discussed in some detail in the following subsections. The user has to choose a method that is appropriate for the application at hand.

All the above mentioned methods are computationally more demanding than ordinary 2D image presentations. The data volumes that are involved to create the images are often large, which requires both fast data transfer and fast computations. There is also usually a minimum frame rate required if the system is to be considered practical by the user. For instance, he or she may want to rotate the data cube in a continuously moving manner to understand which objects are connected etc. A frame rate of less than 10 frames/s is often considered as inconvenient by an operator. Thus, even if the involved computations used in the visualization may be more rudimentary then those used in the reconstruction, the latter may still be computations that require the largest computational power.

Because of this, 3D visualization has earlier only been considered for advanced applications. The cost of professional graphical visualization systems, such as Silicon Graphics work stations, have prevented a use of 3D visualization for many systems. However, in recent years the high competition in the gaming industry has resulted in dramatic improvements in the graphical processing units (GPUs) used in PC's and the performance of the most advanced GPUs is quickly approaching that of professional visualization systems. Several examples of advanced scientific 3D visualization applications using relatively low cost GPUs have been reported lately.

The GPUs can perform highly parallelized computation to speed up these calculations but usually these are performed at a low precision. They also have a relatively limited repertoire. One of the most useful operations is that of calculating the projections of surfaces, and displaying the pixels in textures that are placed on polygon modules in 3D. Such textures are commonly used when creating scenes in computer games. The illusion of semi-transparency can be obtained by combining more than one texture and projecting weighted sums of these textures on the screen. This feature has been used to implement the volume rendering visualization, which is the most advanced mode listed above.

We begin, however, with describing the simplest of these modes, multi-planar rendering, which is an extension of the traditional B-scan mode used both in medicine and NDT.

4.5.1 Multi-planar rendering

In multi-planar rendering, two or more cross-sections of the reconstructed volume are displayed as images attached on planes. If the planes can be moved around quickly and flexibly, this relatively simple display mode can help the inspector in mentally reconstructing the 3D structures. The idea is illustrated in Fig. 4.15 below. The data in the illustration was obtained by simulations of measurements from a 1D phased array acquiring data in the traditional line-by-line fashion with the same focal law in both transmission and reception. The data from the yz-plane was obtained by rotating the array by 90^o.

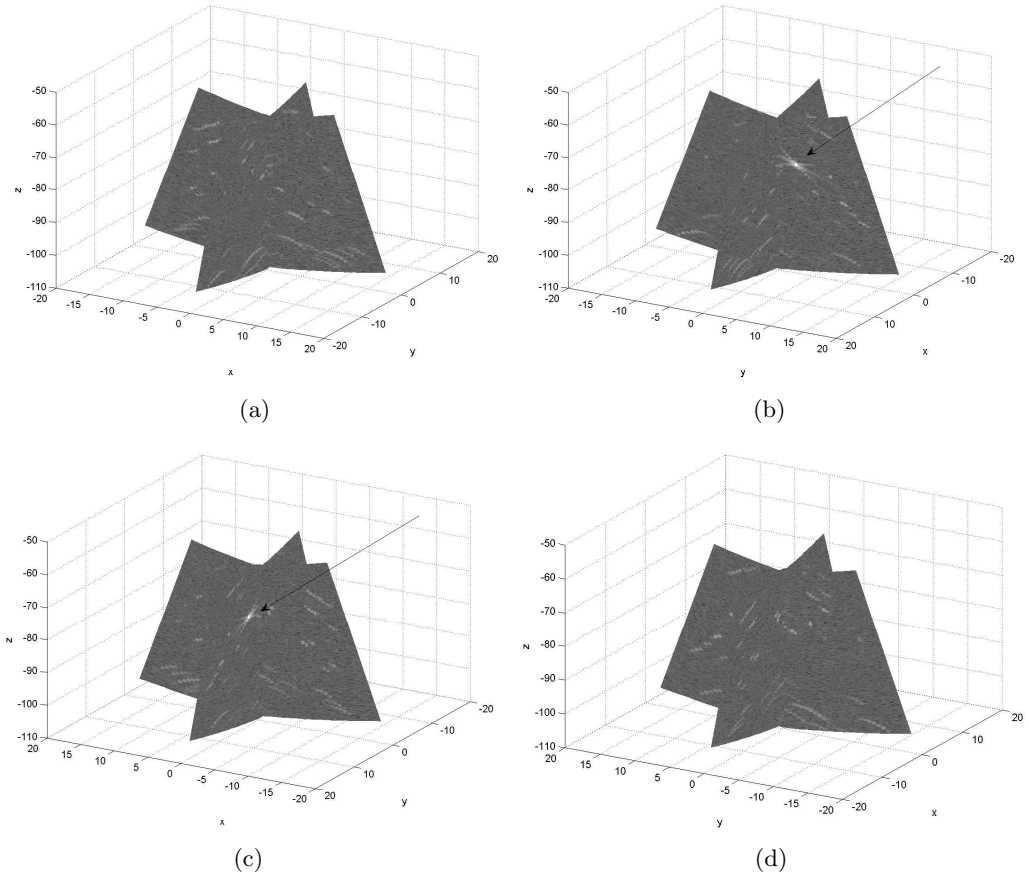


Figure 4.15: Illustration of multi-planar display. Two cross-sections (B-scans) are displayed on two or more planes. The orientation of the planes in the display correspond to the planes that the data were acquired from. Three views of the same data set which was obtained from simulations of a phased array. One strong contribution coming from a point at coordinates $(-2, 1, 80)$ is indicated with arrows in (b) and (c)

Two basic operations are required in multi-planar rendering: (i) Interpolation, in which the values at the grid points are mapped to values at the points on the planes and (ii) projection which maps the 3D positions to the proper point on the screen. The color at the screen are the interpolated intensity values. Note that the interpolation makes possible the visualization of planes in directions other than the scanning directions.

The well developed technique of attaching textures on surfaces mentioned above can be used rather straightforwardly to solve this problem. Since only a small fraction of the data volume is required in the interpolation, the multi-planar visualization method is the fastest of the three mentioned above.

One disadvantage is that the technique still require the operator to mentally create a map of the objects. The connectivity may be difficult to grasp since he or she is able to see a only a very small fraction of the data volume at one time.

4.5.2 Surface based rendering

Surface rendering is an indirect method of obtaining an image from a volume data set. In surface based rendering the data volume is displayed as one or several contour surfaces. The technique relies on the fact that humans are better at perceiving isolated than diffuse objects.

The contours that are shown to the user can either be obtained by a detection procedure or by calculating so-called iso-surfaces which can be seen as the 3D analogue of 2D contour mapping, such as the isobars that are often shown in weather forecasts. If the 3D reconstruction resulted in the function $o(\mathbf{r})$, the iso-surfaces are defined by the equation $o(\mathbf{r}) = k$, where k is a constant.

The data are given at discrete points and interpolation can be used to find the points on the edges between these points that satisfy $o(\mathbf{r}) = k$. The surfaces are built from geometrical primitives, such as polygons or quadratic surfaces, and these primitives are fitted to the crossing points. The obtained iso-surfaces can then be projected into displayable images using standard computer graphics techniques.

Surface rendering suffers from the problem of having to make a binary decision; either a surface passes through a voxel or not. As a consequence, these method often exhibit spurious surfaces or erroneous holes in existing surfaces. This is particularly true for noisy data for which the desired features are difficult to detect and the technique is best suited for imaging well defined object.

In medical imaging, surface based rendering has been used with success only in a few applications. The main application using surface based rendering is imaging of the fetus. Since the baby is surrounded by water, the reflections from the skin are relatively easy to detect. Adding a skin imitating texture on this surface in the display can give a quite realistic impression of the fetus, which helps perceiving details such as facial expressions.

4.5.3 Volume rendering

Instead of mapping geometric primitives onto the data, volume rendering provides a method of directly displaying the data without any intermediate surface representations. It does this by projecting the data directly onto an image screen in a process which involves both viewing and shading the data. This allows it to both retain the details of the internal information on the data, and avoid making any assumptions about the underlying structure of the data.

Volume rendering is based on an optical model, in which the density is thought of as a glowing transluctant gel. If a volume element (voxel) is associated with a large value, the brighter its is, but it then also becomes less transparent. This method of visualizing semi-transparent object is known as *alpha blending*. The intensity that is accumulated along a line is projected on the screen, see Fig. 4.16.

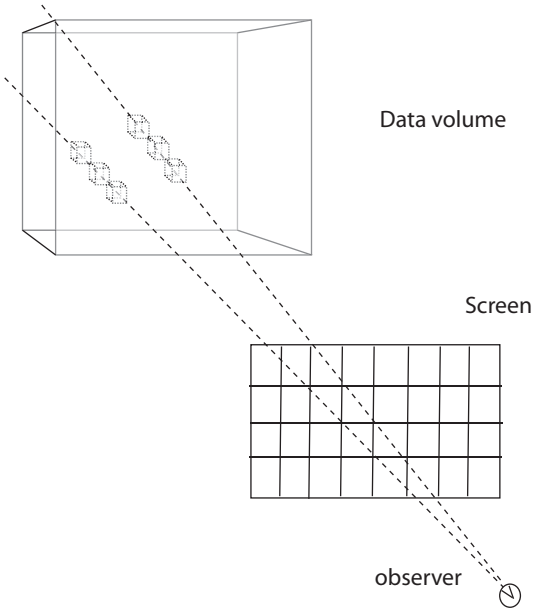


Figure 4.16: Illustration of ray casting used in volume rendering. Each pixel on the screen is associated to a ray (two such are shown in the figure). Each volume element contributes to the values through its "glowing" intensity and its opacity.

For data cubes with very few but very bright voxels, the so-called *maximum intensity projection* can be a suitable simple alternative to the volume rendering. In this method the maximum intensity voxels are located along the ray and only this value is projected.

4.5.4 Hardware for volume rendering

The following GPUs have been reported to have sufficient computing power and flexibility to provide real time volume rendering of relatively large data cubes:

- NVIDIA NV40: In [39] this GPU was tested on data cubes of size $512 \times 512 \times 360$ and for this size the frame rate under interaction was above 10 frames per second (fps). For a smaller data set of size $256 \times 256 \times 109$ the speed was close to 20 fps.
- NVIDIA GeForce 6800.

4.6 3D ultrasonic imaging for FSW inspection

In this final section we briefly point out some aspects that may potentially be important when choosing an appropriate technique for 3D imaging of FSW welds.

There are a number of factors that differ from medical imaging. Some of the most important of these are the following:

- FSW inspection does not involve moving structures. Thus, there are no requirements that the data from a large volume must be acquired at a speed necessary to track any movements in the structure.
- The inspected object is relatively large compared to any realistic array aperture size.

- The scattering object in FSW inspection will have a relatively large density contrast compared to those found in medical applications.
- Wave propagation in solids involve two important wave modes: longitudinal and transversal modes as opposed to the single longitudinal (pressure) mode in medical imaging. These two modes have different propagation speed. The propagation of the longitudinal mode is approximately 4600 m/s which is approximately three times larger than in water. The propagation of the transversal mode that is about half that of the longitudinal, yield about half the wave length for a certain frequency.

The first two factors speak in favor for the mechanical scanning as a practical alternative for the acquisition. A consequence of the third factor is that all parts of the region of interest must be insonified from a fairly large angle interval. Otherwise, there is a risk that some region of interest may lie in the shadow of strong reflectors.

There are a number of consequences of the differences in sound speed and presence of different wave modes mentioned in factor four: first, the presence of a second wave mode complicates somewhat the use of DAS. Focusing laws for both modes would require different delays corresponding to the longitudinal and the transversal sound speed. The second issue concerns analyzing the data due to the transversal mode which can result at the interfaces. The transversal mode that results in a lower propagation speed has the advantage of yielding shorter wavelengths and, thus, better possibilities of obtaining high resolution. However, in most applications, arrays are optimized for longitudinal waves and the transversal mode is normally much weaker than the longitudinal one.

Appendix: Details on pulse compression

Below, the concept of pulse compression mentioned in Section 4.3.2 is illustrated. We consider the case of pulse-echo measurements using a single transducer that have an electro-mechanical impulse responses $h_{em}^{tr}(k)$ and $h_{em}^{rec}(k)$ in transmit and receive, respectively. The different steps can be described mathematically as follows: The transmitted pulse, $p_{tr}(k)$, is the convolution between the excitation (pulse compression) signal, $x(k)$, and $h_{em}^{tr}(k)$

$$p_{tr}(k) = x(k) * h_{em}^{tr}(k). \tag{4.18}$$

We assume here that the number of sample in $x(k)$ is L.

The back-propagating echo from a point scatterer, $p_{rec}(k)$, is a replica of $p_{tr}(k)$ delayed with the two-way travel time between the transducer and the scatterer, Δ, and scaled with an amplitude o. The electrically acquired signal is modeled as a convolution between this incoming wave and $h_{em}^{rec}(k)$:

$$p_{rec}(k) = o \cdot p_{tr}(k) * h_{em}^{rec}(k) * \delta(k - \Delta). \tag{4.19}$$

The output is obtained by applying a matched filter to the received signal,

$$y_x(k) = p_{rec}(k) * x(L - k) = o \cdot h_{em}(k) * c_x(k) * \delta(k - \Delta). \tag{4.20}$$

where $c_x(k) \triangleq x(k) * x(L - k)$ is the autocorrelation function, delayed L samples, and $h_{em}(k) = h_{em}^{tr}(k) * h_{em}^{rec}(k)$.

The signal $y_x(k)$ in (4.20) should be compared with the signal, $y_0(k)$, that we obtain using an excitation that is the standard impulse at $k = 0$ with an amplitude a_{exc}:

$$y_0(k) = h_{em}^{tr}(k) * h_{em}^{rec}(k) * x(k) * x(L - k) * \delta(k - \Delta) = o \cdot a_{exc} h_{em}(k) * \delta(k - \Delta). \tag{4.21}$$

The expressions in (4.21) and (4.20) differ by a convolution factor $c_x(k)$ and if $c_x(k)$ can be well approximated by a delta function, the signals $y_x(k)$ and $y_0(k)$ will differ only in the time delay L and a scaling factor. So, obviously, if the goal is to simulate signals obtained with a strong impulsive, $x(k)$ should be chosen so that $c_x(k)$ closely resembles a delta function.

A common example of a pulse compression signal is the linear frequency modulated (FM) sinusoid [40–43], often called the *chirp* signal. Fig. 4.17 illustrates pulse compression steps in eqs. (4.18)-(4.20) with $x(k)$ being a chirp signal. We note that the waveform of the output from the matched filter, shown in Fig. 4.17(e), fairly well resembles the signal simulated under the assumption of impulsive excitation shown in Fig. 4.17(f).

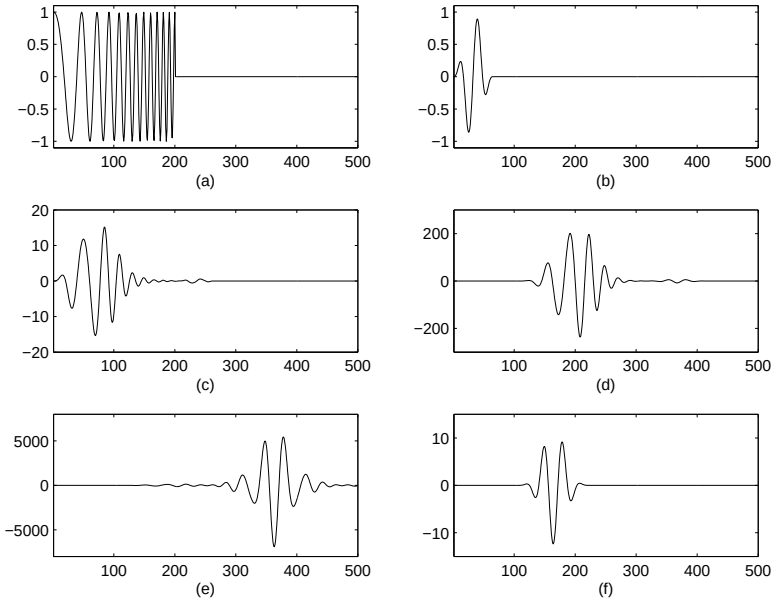


Figure 4.17: Illustration of pulse compression using chirp excitation. (a) A linear chirp of length $L = 200$ is used for exciting the transducer element. This element has an electro-mechanical impulse response shown in (b). (c) The transmitted pulse is the convolution between the chirp and $h(k)$. (d) The received signal is a the transmitted pulse convolved with the backward electro-mechanical impulse response. (e) the output from a matched filter applied to the received signal. (f) The signal that would be received with an impulsive excitation. The main differences between the signals in (e) and (f) are the strength and a time shift of $L = 200$.

In [26], Misaridis and Jensen compared a number of alternative excitation signals for pulse compression. Their conclusion was that chirp signals are the most robust to the frequency dependent attenuation that is present in most practical ultrasonic imaging applications.

There are however several important alternatives to chirp signals. Examples of such are so-called Barker codes which are bipolar sequences $b_i = +/-1$ of length $L \geq 2$ such that the autocorrelation sequence $c_b(l) \triangleq \sum_i b_i b_{i+l}$ fulfils $c_b(l) = L$ for $l = 0$ and $|c_b(l)| \leq 1$ for $l \neq 0$. Below, one of the Barker codes of length 13 is displayed with its corresponding autocorrelation function.

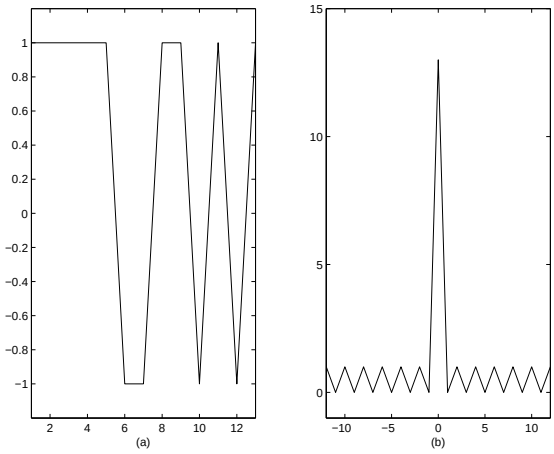


Figure 4.18: One of the Barker codes of length 13 (a) and its autocorrelation function (b).

Bipolar sequences are attractive because they are robust to static non-linearities in the amplifier used at transmission [44]. Unfortunately there are known Barker only up to length 13. Therefore, these are not good candidates if very long signals are required.

No finite time excitation signal has the property that $c_x(k) = K\delta(k-L)$, i.e., it is impossible to exactly simulate the behavior of a strong impulsive excitation. However, it is possible if we allow using two transmissions in the compression. The technique relies on so-called *Golay codes* [44, 45] and is described below.

Golay codes

Golay codes come in pairs that have the property that their autocorrelation functions sum to a delta function. Let $g_1(k)$ and $g_2(k)$ be two bipolar sequences of length L and let $c_1(k)$ and $c_2(k)$ denote their respective autocorrelation sequences defined as

$$c_i(l) = \sum_{l=1}^{L} g_i(l-k)g_i(k). \tag{4.22}$$

A Golay pair, $g_1(k)$ and $g_2(k)$, has the property that

$$c_1(l) + c_2(l) = \begin{cases} 2L & \text{if } l = 0 \\ 0 & \text{else.} \end{cases} \tag{4.23}$$

This property can be used for distortion-less pulse compression as follows: Transmit at different times two separate pulses using $g_1(k)$ and $g_2(k)$ for the excitation. The output from applying matched filters at received signals from each reception can be written as (cf. eq. (4.20))

$$y_i(k) = h_{em}(k) * c_i(k) * o \cdot \delta(k-\Delta) \tag{4.24}$$

and by combining (4.23) and (4.24), we obtain the sum

$$y(k) = y_1(k) + y_2(k) = h_{em}(k) * (c_1(k) + c_2(k)) * o \cdot \delta(k-\Delta) = L \cdot o \cdot h_{em}(k) * \delta(k-\Delta) \tag{4.25}$$

In contrast to the pulse compression described earlier, the obtained signal $y(k)$ is exactly a scaled and time-delayed replica of $y_0(k)$.

Bibliography

[1] A. Fenster and D.B. Downey. Three-dimensional ultrasound imaging. *Annu. Rev. Biomed. Eng.*, 114(1):457–475, July 2000.

[2] S. Tezuka, S. Hashimoto, T. Togashi, Y. Miyajima, and Y. Seo. A two-dimensional array probe that has a huge number of active channels. In *2003 IEEE Ultrasonics Symposium*, volume 1, pages 960–963, 2003.

[3] D.P. Shattuck and M.D. Weinshenker. Explososcan: A parallel procesing technique for high speed ultrasound imaging with linear phased arrays. *J. Acoust. Soc. Am.*, 75(4):1273–1282, April 1984.

[4] D.G. Wildes, R.Y. Chiao, C.M.W Daft, K.W. Rigby, L.S. Smith, and K.E. Thomenius. Elevation performance of 1.25d and 1.5d transducer arrays. *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, 44(5):1027 – 1037, 1997.

[5] O.T. von Ramm, S.W. Smith, and F.L Thurstone. Grey scale imaging with complex tgc and transducer arrays. In *Proc. Soc. Photo-opt Inst. Engineers, Medicine IV*, volume 70, pages 226–270, 1975.

[6] S.W. Smith, H.G. Pavy, and O.T. von Ramm. High-speed ultrasound volumetric imaging system–part I: Transducer design and beam steering. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 38(2):100–108, March 1991.

[7] O.T. von Ramm, S.W. Smith, and H.G. Pavy. High-speed ultrasound volumetric imaging system–part I: Parallel processing and image display. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 38(2):109–115, March 1991.

[8] S.M. Gehlbach and R.E. Alvarez. Digital ultrasound imaging techniques using vector sampling and raster line reconstruction. *Ultrasonic Imaging*, 3(1):83–107, January 1981.

[9] G.R. Lockwood and F.S. Foster. Optimizing sparse two-dimensional transducer arrays using an effective aperture approach. In *1994 IEEE Ultrasonics Symposium*, volume 3, pages 1497–1501, 1995.

[10] B. Steinberg. Comparison between the peak sidelobe of the random array and algorithmically designed aperiodic arrays. *IEEE Trans. Antennas and Propagation*, 21(3):366–370, May 1973.

[11] G.R Lockwood, P-C. Li, and M.D. O'Donnell. Optimizing the radiation pattern of sparse periodic linear arrays. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 43(1):7–14, January 1996.

[12] C.B. Burckhardt, P.A. Grandchamp, and H. Hoffmann. Experimental 2 Mhz synthetic aperture sonar system intended for medical use. *IEEE Transactions on Sonics and Ultrasonics*, SU-21(1):1 – 6, 1974.

[13] W.R. Fenner and G.E. Stewart. Synthetic aperture holographic imaging a medical application. *Metals Technology*, pages 75 – 79, 1975.

[14] R. C. Fairchild, B. H. Anderson, and J. R. Frederick. Synthetic aperture ultrasonic inspection of pressure vessels and piping. *American Society of Mechanical Engineers (Paper)*, (77-PVP-23):8 –, 1977.

[15] G.E. Trahey and L.F. Nock. Synthetic receive aperture imaging with phase correction for motion and for tissue inhomogeneities- Part I: Basic principles. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 39(4):489–495, July 1992.

[16] G.E. Trahey and L.F. Nock. Synthetic receive aperture imaging with phase correction for mation and for tissue inhomogeneities- Part II: Effects of and correction for motion. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 39(4):496–501, July 1992.

[17] G.R Lockwood, J.R. Talman, and S.S Brunke. Real-time 3-d ultrasound imaging using sparse synthetic aperture beamforming. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 45(4):980–988, July 1998.

[18] S.I. Nikolov. *Synthetic Aperture tissue and flow ultrasound imaging.* PhD thesis, Technical University of Denmark, August 2001.

[19] J. Ylitalo, E. Alasaarela, and J. Koivukangas. Ultrasound holographic b-scan imaging. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 36(3):376–383, May 1989.

[20] M. Karaman, P.C. Li, and M. O'Donnell. Synthetic aperture imaging for small scale systems. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 42(3):429–442, May 1995.

[21] G.R. Lockwood and F.S. Foster. Design of sparse array imaging systems. In *1995 IEEE Ultrasonics Symposium*, volume 2, pages 1237–1243, 1995.

[22] C. Passmann and H. Ermert. In vivo imaging of the skin in the 100 Mhz region using the synthetic aperture concept. In *1995 IEEE Ultrasonics Symposium*, volume 2, pages 1287 – 1290, 1995.

[23] M.-H. Bae and M.-K. Jeong. A study of synthetic-aperture imaging with virtual source elements in b-mode ultrasound imaging systems. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 47(6):1510–1519, Nov 2000.

[24] E. Wennerström and T. Stepinski. Model based correction of diffraction effects of the virtual source element. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 2006.

[25] R.Y. Chiao, L.J. Thomas, and S-D Silverstein. Sparse array imaging with spatially-encoded transmits. In *1997 IEEE Ultrasonics Symposium*, volume 2, pages 1679–1682, 1997.

[26] T.X. Misaridis and J.A. Jensen. Use of modulated exciation signals in medical ultrasound. part I: Basic concepts and expected benefits. 52(2):177–191, February 2005.

[27] S.J. Norton and M. Linzer. Ultrasonic reflectivity imaging in three dimensions: Exact inverse scattering solutions for plane, cylindrical, and spherical apertures. *IEEE Trans. Biomed. Eng.*, 28(2):202–220, Feb 1981.

[28] T.D Mast. Wideband quantitative ultrasonic imaging by time-domain diffraction tomography. *J. Acoust. Soc. Am.*, 106(6):3061–3071, December 1999.

[29] David Nahamoo, S. X. Pan, and Avinash C. Kak. Synthetic aperture diffraction tomography and its interpolation-free computer implementation. *IEEE Trans. Sonics Ultras.*, SU-31(4):218 – 229, 1984.

[30] S. Kenue and J.F. Greenleaf. Limited angle multifrequency diffraction tomography. *IEEE Trans. Sonics Ultras.*, SU-29(4):213 – 217, 1982.

[31] A. Yamada and S. Yano. Ultrasound inverse scattering computed tomography under the angular illumination limitation. *Japanese Journ. Appl. Phys., Part 1: Regular Papers and Short Notes and Review Papers*, 43(8 A):5582 – 5588, 2004.

[32] K.J. Langenberg, M. Berger, T. Kreutter, and K. Mayer. Synthetic aperture focusing technique signal processing. *NDT International*, 19(3):177–189, June 1986.

[33] J. Lu and J.F. Greenleaf. Ultrasonic nondiffracting transducer for medical imaging. *IEEE Trans. Ultrason., Ferroelect., and Freq. Contr.*, 37.

[34] J. Lu and J.F. Greenleaf. Experimental verification of nondiffracting x-waves. *IEEE Trans. Ultrason., Ferroelect., and Freq. Contr.*, 39(3):441–446, May 1992.

[35] H.F. Choi, J. D'Hooge, B. Bijnens, W. Lauriks, J. Thoen, and P. Suetens. Minimal transducer requirements to implement limited diffraction beams in clinical practice. In *2002 IEEE Ultrasonics Symposium*, volume 2, pages 1605 – 1608, 2002.

[36] J. Lu. 2d and 3d high frame rate imaging with limited diffraction beams. *IEEE Trans. Ultrason., Ferroelect., and Freq. Contr.*, 44(4):839 – 856, July 1997.

[37] J. Lu. Experimental study of high frame rate imaging with limited diffraction beams. *IEEE Trans. Ultrason., Ferroelect., and Freq. Contr.*, 45(1):84 – 97, January 1998.

[38] M. Levoy. Display of surfaces from volume data. *IEEE Computer Graphics and Applications*, 8(3):29–37, May 1988.

[39] Y. Heng and L. Gu. Gpu-based volume rendering for medical image visualization. In *Proc. 2005 IEEE Eng. in Med. Biol. 27th Ann. Conf.*

[40] M. O'Donnell. Coded excitation system for improving the penetration of real-time phase.array imaging systems. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 39(3):341–351, May 1992.

[41] P. Li, E. Ebbini, and M. O'Donnell. A new filter design technique for coded excitation systems. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 39(6):693–699, Nov 1992.

[42] M. Pollakowski and H. Ermert. Chirp signal matching and signal power optimization in pulse-echo mode ultrasonic nondestructive testing. *IEEE Trans. Ultrason., Ferroelect. and Freq. Contr.*, 41(5):655–659, September 1994.

[43] T.X. Misaridis and J.A. Jensen. An effective decoded excitation scheme based on a predistorted FM signal and an optimized digital filter. In *2003 IEEE Ultrasonics Symposium*, 1999.

[44] M.H. Bae, W.Y. Lee, M.K. Jeong, and S.J. Kwon. Orthogonal Golay code based ultrasonic imaging without reducing frame rate. In *2002 IEEE Ultrasonics Symposium*, 2002.

[45] R.Y. Chiao and L.J. Thomas. Synthetic transmit aperture imaging using orthogonal Golay coded excitation. In *2000 IEEE Ultrasonics Symposium*, volume 2, pages 1667–1680, 2000.