Differentiable Beamforming
for Ultrasound Autofocusing

Walter Simson$^{\star 1[0000-0002-2801-8646]}$, Louise Zhuang$^{1[0000-0002-5677-4895]}$,
Sergio J. Sanabria$^{1[0000-0003-4786-4597]}$, Neha Antil$^{1[0000-0001-7896-9231]}$,
Jeremy J. Dahl$^{1[0000-0001-9877-452X]}$, and
Dongwoon Hyun$^{1[0000-0003-2625-8109]}$

Stanford University, Stanford, CA 94305 USA
{waltersimson, dongwoon.hyun}@stanford.edu

Abstract. Ultrasound images are distorted by phase aberration arising from local sound speed variations in the tissue, which lead to inaccurate time delays in beamforming and loss of image focus. Whereas state-of-the-art correction approaches rely on simplified physical models (e.g. phase screens), we propose a novel physics-based framework called differentiable beamforming that can be used to rapidly solve a wide range of imaging problems. We demonstrate the generalizability of differentiable beamforming by optimizing the spatial sound speed distribution in a heterogeneous imaging domain to achieve ultrasound autofocusing using a variety of physical constraints based on phase shift minimization, speckle brightness, and coherence maximization. The proposed method corrects for the effects of phase aberration in both simulation and in-vivo cases by improving image focus while simultaneously providing quantitative speed-of-sound distributions for tissue diagnostics, with accuracy improvements with respect to previously published baselines. Finally, we provide a broader discussion of applications of differentiable beamforming in other ultrasound domains.

Keywords: Ultrasound · Image reconstruction · Optimization.

1 Introduction

Ultrasound images are reconstructed by time sampling the reflected pressure signals measured by individual transducer elements in order to focus at specific spatial locations. The sample times are calculated so as to compensate for the time-of-flight from the elements to the desired spatial locations, often by assuming a constant speed of sound (SoS) in the medium, e.g., 1540 m/s. However, the human body is highly heterogeneous, with slower SoS in adipose layers than in fibrous and muscular tissues. If unaccounted for, these differences lead to phase aberration, geometric distortions, and loss of focus and contrast [1]. This degradation is a fundamental limitation of current ultrasound image reconstruction and impacts downstream tasks such as diagnostics, volumetry, and registration.

---

* corresponding author

Historically, phase aberration has been described using simplified phase-screen models [19, 5], which assume that distortions generated from an unknown SoS can be modeled by a gross time delay offset at every element [1]. More recently, several methods have been proposed to estimate SoS distribution of the medium from aberration measurements as a step before actual image correction. A family of these methods still relies on simplified physical models of wave propagation to derive tractable inverse problems. These include assuming a horizontally layered medium [9] or coherent plane wavefront propagation at different angulations [17]. To reinforce specific assumptions about SoS heterogeneity, regularization is often introduced, including total variation for focal inclusion geometries [15] and Tikhonov regularization for smoothly varying layered SoS distributions [17, 14]. While these methods perform well for one class of SoS inversion problems, it is challenging to generalize their applicability to arbitrary SoS distributions, which are generally found in clinical scenarios. Work has been carried out to find more generalizable estimation based on training neural network models to end-to-end learn SoS distributions or optimize the regularization function basis [18, 4, 16]. However, these methods require thousands of training instances, which can currently practically only be obtained from in-silico simulations and show challenges generalizing to real data.

Recent developments in artificial intelligence have been facilitated by the release of open-source tensor libraries, which can perform automatic differentiation of composable transformations on vector data. These libraries are the backbone of complex neural network architectures that use automatic reverse-mode differentiation (back-propagation) to iteratively optimize weights based on a set of training instances. These libraries also simplify and optimize portability to high-performance computing platforms. We hypothesize that such libraries can likewise be extended to model the pipeline of ultrasound image reconstruction as a composition of differentiable operations, allowing optimization based on a single data instance.

In this work, we propose an ultrasound imaging paradigm that jointly achieves sound speed estimation and image quality enhancement via differentiable beamforming. We formulate image reconstruction as a differentiable function of a spatially heterogeneous SoS map, and optimize it based on quality metrics extracted from the final reconstructed images (Fig. 1).

2   Methods

2.1   Beamforming Multistatic Synthetic Aperture Data

In ultrasound imaging, radiofrequency data (RF) represents the time series signal proportional to the pressure measured by each probe array sensor. A multistatic synthetic aperture dataset contains the RF pulse-echo responses of every pair of transmit and receive elements. We denote the signal due to the i-th transmit element, and j-th receive element as $u_{ij}(t)$. This signal can be focused to an arbitrary spatial location $\mathbf{x}_k$ by sampling $u_{ij}(t)$ at the time corresponding to the
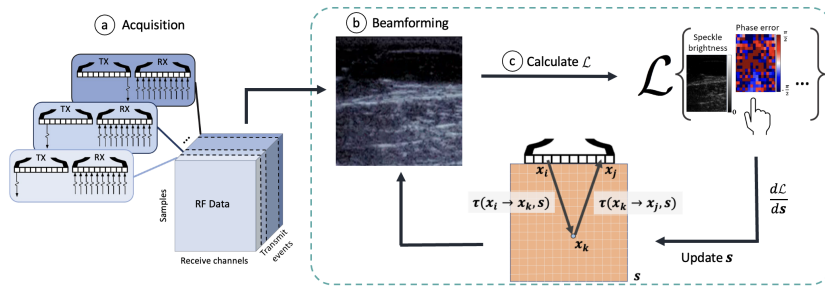
Fig. 1. Differentiable beamforming method for ultrasound autofocusing. Part (a) shows the initial full synthetic aperture data acquisition. The complete RF data is then used for beamforming in part (b) with an initial estimate of slowness, and afterwards, a desired loss is calculated in part (c). The loss is differentiated with respect to the slowness, which is then updated and used for the next iteration of beamforming. This process encapsulated in the box is then repeated until convergence is reached.

time-of-flight τ from the transmit element at $\mathbf{x}_i$ to $\mathbf{x}_k$ and back to the receive element at $\mathbf{x}_j$, achieved via 1D interpolation of the RF signal:

$$u_{ij}(\mathbf{x}_k) = u_{ij}\left(\tau(\mathbf{x}_i, \mathbf{x}_k) + \tau(\mathbf{x}_k, \mathbf{x}_j)\right). \tag{1}$$

(We describe our time-of-flight model in greater detail below in Sec. 2.3.) The interpolated signals are then summed across the transmit (N_t) and receive (N_r) apertures to obtain a focused ultrasound image:

$$u(\mathbf{x}_k) = \sum_{i=1}^{N_t} \sum_{j=1}^{N_r} u_{ij}(\mathbf{x}_k). \tag{2}$$

This process of interpolation and summation is called delay-and-sum (DAS) beamforming.

2.2   Differentiable Beamforming

DAS is composed of elementary differentiable operations and is consequently itself differentiable. Therefore, DAS can be incorporated into an automatic differentiation (AD) framework to allow for differentiation with respect to any desired input parameters $\boldsymbol{\theta}$. For a given loss function $\mathcal{L}(u(\mathbf{x}_k; \boldsymbol{\theta}))$ that measures the "quality" of the beamforming, $\boldsymbol{\theta}$ can be optimized using gradient descent to identify the optimal $\boldsymbol{\theta}^\star$ using update steps $\Delta\boldsymbol{\theta}$:

$$\boldsymbol{\theta}^\star = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(u(\mathbf{x}_k; \boldsymbol{\theta})), \qquad \Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \alpha \frac{\partial}{\partial\boldsymbol{\theta}} \mathcal{L}(u(\mathbf{x}_k; \boldsymbol{\theta})). \tag{3}$$

This differentiable framework is flexible, providing many ways to parameterize the beamforming. In this work, we will show the promise of differentiable beamforming on the task of sound speed estimation by optimizing for slowness $\mathbf{s}$ in a time of flight delay model (i.e. $\boldsymbol{\theta} = \mathbf{s}$).

2.3   Time of Flight Model

Here, we parameterize the slowness (i.e. the reciprocal of the sound speed) as a function of space. Specifically, we define the slowness at a set of control points as $\mathbf{s} = \{s(\mathbf{x}_k)\}_k$, which can be interpolated to obtain the slowness at arbitrary $\mathbf{x}$. The time-of-flight from $\mathbf{x}_1$ to $\mathbf{x}_2$ is the integral of the slowness along the path:

$$\tau(\mathbf{x}_1, \mathbf{x}_2; \mathbf{s}) = \int_{\mathbf{x}_1 \to \mathbf{x}_2} \mathbf{s} \, d\mathbf{x}. \tag{4}$$

For simplicity and direct comparison with previous sound speed estimation models [17], a straight ray model of wave propagation is used.

2.4   Loss Functions for Sound Speed Optimization

Speckle Brightness Maximization Diffuse ultrasound scattering produces an image texture called speckle. Speckle brightness can be used as a criterion of focus quality [13]. Written as a loss, this is the negative average pixel magnitude:

$$\mathrm{SB}(\mathbf{s}) = \frac{1}{N_k} \sum_k |u(\mathbf{x}_k; \mathbf{s})| = -\mathcal{L}_{\mathrm{SB}}(\mathbf{s}). \tag{5}$$

Coherence Factor Maximization Coherence factor [6, 11], also referred to as the F criterion or "focusing criterion", defined between 0 and 1, is the measure of the coherent signal sum over the incoherent signal sum of the receive aperture. When received signals are in focus (i.e. in equal phase), CF achieves the maximum value of 1. We use the negative CF as a loss:

$$\mathrm{CF}(\mathbf{s}) = \frac{1}{N_k} \sum_{k=1}^{N_k} \frac{\left| \sum_j \sum_i u_{ij}(\mathbf{x}_k; \mathbf{s}) \right|}{\sum_j \left| \sum_i u_{ij}(\mathbf{x}_k; \mathbf{s}) \right|} = -\mathcal{L}_{\mathrm{CF}}(\mathbf{s}). \tag{6}$$

Phase-Error Minimization The van Cittert Zernike theorem of optics [12] states that when imaging diffuse scatterers using a given transmit and receive sub-aperture $\mathcal{T}_a$ and $\mathcal{R}_a$ (i.e. subset of the available array elements), the resulting signal is almost perfectly correlated with the signal from a second set of apertures $\mathcal{T}_b$ and $\mathcal{R}_b$ when the two apertures share a common midpoint. The measured phase-shift between both signals should approach zero when aberration is corrected. Fig. 2 illustrates this concept of phase error.

We estimate the phase shift as the complex angle between DAS signals u_a and u_b of the respective subapertures $(\mathcal{T}_a, \mathcal{R}_a)$ and $(\mathcal{T}_b, \mathcal{R}_b)$, calculated using (2):

$$\Delta\phi_{ab}(\mathbf{x_k}) = \angle \mathbb{E}[u_a(\mathbf{x_k}; \mathbf{s}) u_b^*(\mathbf{x_k}; \mathbf{s})]. \tag{7}$$

The phase shift error (PE) is defined for a set of all aperture pairs (a, b) with common midpoint as

$$\mathrm{PE}(\mathbf{s}) = \frac{1}{N_{(a,b)}} \sum_{(a,b)} |\Delta\phi_{ab}| = \mathcal{L}_{\mathrm{PE}}(\mathbf{s}). \tag{8}$$
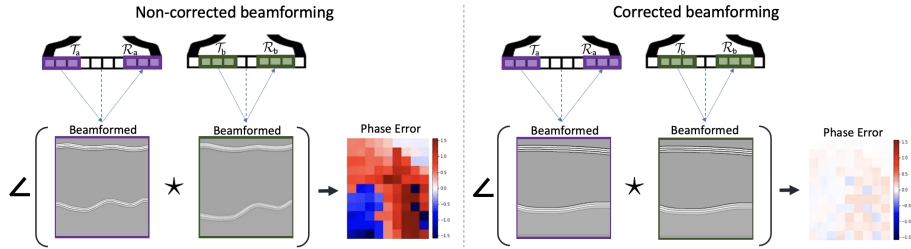
Fig. 2. Phase error minimization in correlated common mid-point sub-apertures. Phase error is computed as the angle of the cross correlation of complex beamformed signals from different sub-apertures sharing a common midpoint. When the correct slowness is used for the beamforming, the phase error is minimized.

3 Experimental Setup

3.1 Implementation of Differentiable Beamformer

A differentiable DAS beamformer was implemented in Python using JAX[1] [3], which provides out of the box GPU acceleration. DAS was parameterized by the slowness map, where the time-of-flights for beamforming were calculated via bilinear interpolation of the slowness along a discretized path from the transmitting element to a location of interest and from the location to a receiving element. The loss was computed on 5×5 pixel patches ($\lambda/2$ pixel spacing) on a regular 15×21 grid spanning the image. The sound speed map was then optimized via gradient descent. For the phase error loss, 17-element subapertures were used for beamforming. The beamformed data for every subaperture pair with a common midpoint were cross-correlated with a 5×5 path to compute the phase shift. We further leveraged acoustic reciprocity to combine the results for reciprocal transmit/receive subapertures. This phase-shift measurement was then used for the final phase error loss. The GPU-based implementation runs in $\sim$300 seconds for 300 iterations on an NVIDIA RTX A6000. The code for this work can be found on GitHub[2].

3.2 Comparison with state-of-the-art methods

As a baseline for performance comparison, the Computed Ultrasound Tomography in Echo Mode (CUTE) method developed by Stähli et al. [17] was implemented in MATLAB; this method has been shown to achieve sound speed reconstruction of both layered and focal lesion geometries. The method shows some similarities in using phase error minimization from different apertures (albeit in the angular domain) and ray tracing paths. However, it relies on a coherent plane wavefront propagation model and Tikhonov regularization to build a tractable inverse problem.

---

[1] https://github.com/google/jax
[2] https://github.com/waltsims/dbua

3.3   Datasets

In-silico The CUDA-accelerated binaries of the k-Wave simulation suite [10] were used to generate multistatic RF data of 3D phantom model acquisitions. To compare with the baseline [17], simulations were first generated using plane-wave transmissions (115 transmits in steering range of -28.5°:0.5°:28.5°) and then converted to FSA format using REFoCUS [2] in the rtbf framework [7]. A linear 128 element linear probe was simulated, with a pitch of 0.3 mm and a center frequency of 4.8 MHz with a 100% bandwidth. The simulation domain was 60 x 51 x 7.4 mm^3. Iso-echoic phantoms were generated whereby the sound speed was modulated relative to the density of a region so the average brightness remained constant while the sound-speed variation introduced phase aberration.

In-vivo In-vivo data was collected on a Verasonics Vantage research system with a L12-3v linear transducer (192 elements, 0.2 mm pitch, 5 MHz center frequency). Three abdominal liver views, which contained subcutaneous adipose, musculoskeletal tissue and liver parenchyma, were collected from a healthy volunteer under a protocol approved by an institutional review board.

4   Results

Figure 3 shows SoS maps for in-silico phantom data. In the uncorrected (naive) B-modes, regions of darkening and smeared speckle can be seen as acoustic intensity diminishes due to aberration. In the quadrant phantom (a), a distinct spatial skewing can be observed from left to right. On the corrected images, image brightness is enhanced, iso-echogenic speckle distributions are revealed, aberrated regions are reduced, and the boundary between quadrants shows a congruent left-to-right and top-to-bottom transition. Similarly, in the inclusion phantom (b), characteristic triangles can be seen to the left and right of the inclusion in the naive B-mode. These triangular offshoots are artifacts produced by total wave reflection on the lateral lesion boundaries when the ultrasound wave encounters an SoS transition at grazing incidence. Moreover, diffraction of waves through the lesion lead to aberration errors behind the lesion. In the corrected B-mode, these dark regions are enhanced, and the image has an overall more homogeneous brightness pattern.

The sound speed distributions generated with differentiable beamforming are in general agreement with the ground truth sound speed distributions. Table 1 quantitatively compares the mean absolute error (MAE) and standard deviation (std) with respect to the ground truth. For all phantoms, differential beamforming achieved lower (better) error metrics than the baseline. Homogeneous phantoms were best reconstructed via CF loss function, while inhomogeneous phantoms were best reconstructed via PE loss function.

Figure 4 shows preliminary results with differential beamforming for the reconstructed in-vivo data. The SoS reconstruction successfully delineates abdominal layers including subcutaneous adipose fat (average 1494 m/s), muscle
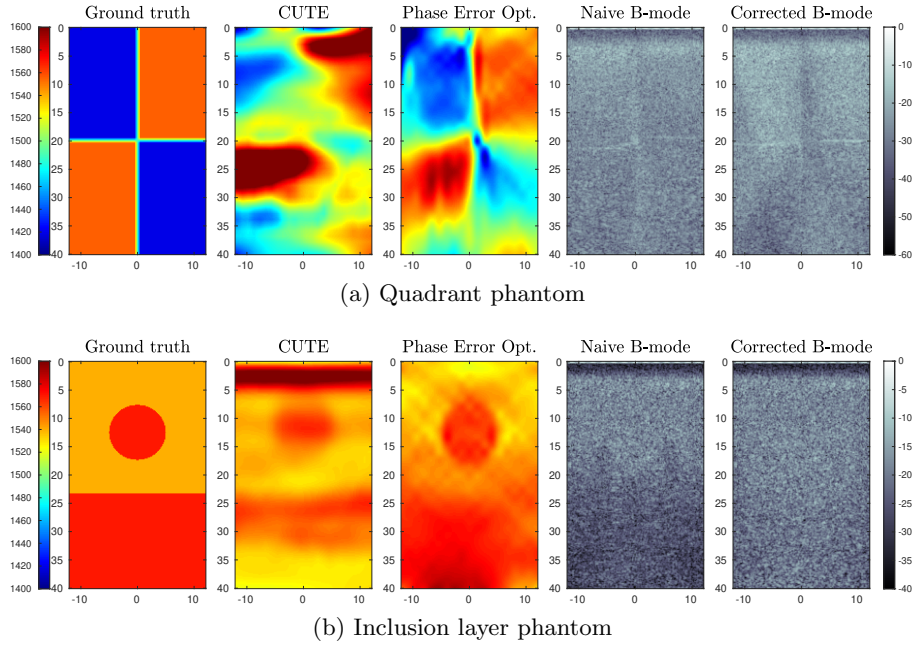
(a) Quadrant phantom

(b) Inclusion layer phantom

Fig. 3. The results of the imaging technique are shown. From left to right, each row shows: 1) the ground truth sound speed; 2) the CUTE method as a baseline; 3) our proposed phase error optimization; 4) a naive B-mode image formed assuming 1540 m/s; and 5) the B-mode reconstructed according to our proposed sound speed estimates. (a) The geometric distortion at the tissue interfaces is corrected. (b) The B-mode image brightness becomes more homogeneous in the lower half of the image. Videos are provided in the supplementary material.

Phantom	Description	CUTE (baseline)	Speckle Brightness	Coherence Factor	Phase Error (proposed)
1420	homogenous	21.6±21.4	3.9±3.3	**3.2±2.6**	4.8±3.5
1465	homogenous	11.7±18.8	4.5±4.9	5.3±4.6	**4.5±3.5**
1480	homogenous	10.4±18.5	6.1±5.4	**4.1±4.2**	4.7±3.5
1510	homogenous	10.8±17.0	6.1±7.0	**4.4±4.5**	4.8±3.6
1540	homogenous	11.8±15.8	7.8±7.5	**5.1±4.4**	6.1±4.3
1555	homogenous	11.4±15.3	**5.7±6.7**	5.8±4.7	5.9±4.3
1570	homogenous	11.2±14.8	7.5±7.6	**4.9±4.7**	6.5±4.7
Quadrant	Fig.3a	65.6±36.3	63.2±52.1	63.4±47.7	**35.4±27.9**
Two layer	[17]	40.2±34.1	62.5±54.2	33.2±25.8	**13.4±14.7**
Four layer	[17]	44.1±27.5	50.5±25.0	43.8±23.2	**29.0±26.5**
Inclusion	[17]	14.3±16.4	8.3±7.5	7.5±5.9	**6.1±4.4**
Inclusion layer	[17], Fig.3b	19.8±18.1	16.3±14.8	15.0±11.1	**7.5±5.0**

Table 1. Comparison of sound speed mean absolute error (MAE) ± standard error between state-of-the-art (CUTE) versus differential beamforming with speckle brightness, coherence factor, and phase error objective functions. Two layer, four layer, inclusion and inclusion layer definitions can be found in [17]. (figure 6a to 6e)

(average 1551 m/s) and liver parenchyma (average 1530 m/s) in agreement with the literature values [8].
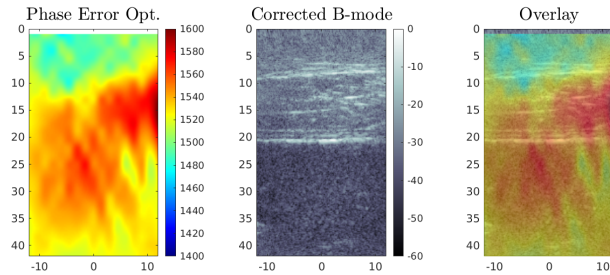

Fig. 4. A sample of in-vivo data reconstructed with the estimated sound speed via differentiable beamforming. Three layers consisting of subcutaneous adipose fat, muscle, and liver parenchyma are visible from top to bottom.

5   Discussion and Conclusion

Differentiable beamforming can be used to solve for unknown quantities with gradient descent. Here, we parameterized beamforming as a function of the slowness and optimized with respect to several candidate loss functions, showing that phase error was best for heterogeneous targets. The differentiable beamformer simultaneously provided B-mode image correction and quantitative sound speed characterization beyond the state-of-the-art across several challenging cases. Preliminary in-vivo quantitative SoS data for liver was shown, which has direct clinical applications such as in the noninvasive assessment of non-alcoholic fatty liver disease, as well as image enhancement in general.

Importantly, the differentiable beamformer allows us to incorporate fundamental physics principles like wave propagation, reducing the number of parameters to optimize. In the future, more complex wave propagation physics, such as refraction models, can be added to SoS optimization. In addition to sound speed, this work can be readily adapted to a broad set of applications such as beamforming with flexible arrays, where element positions are unknown, or passive cavitation mapping, where the origin of the signal is uncertain. Because the gradients flow through the entire imaging pipeline, the differentiable beamformer is also highly compatible with deep learning techniques. For instance, a model can be trained in a self-supervised fashion to identify optimal sound speed updates to accelerate convergence. Differentiable beamforming also enables the end-to-end optimization of imaging parameters for downstream tasks in computer-aided medical diagnostics.

References

1. Ali, R., Brevett, T., Zhuang, L., Bendjador, H., Podkowa, A.S., Hsieh, S.S., Simson, W., Sanabria, S.J., Herickhoff, C.D., Dahl, J.J.: Aberration correction in diagnostic ultrasound: A review of the prior field and current directions. Zeitschrift für Medizinische Physik (2023). https://doi.org/10.1016/j.zemedi.2023.01.003
2. Ali, R., Herickhoff, C.D., Hyun, D., Dahl, J.J., Bottenus, N.: Extending retrospective encoding for robust recovery of the multistatic data set. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **67**(5), 943–956 (2019)
3. Bradbury, J., Frostig, R., Hawkins, P., Johnson, M.J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., Zhang, Q.: JAX: composable transformations of Python+NumPy programs (2018), http://github.com/google/jax
4. Feigin, M., Freedman, D., Anthony, B.W.: A deep learning framework for single-sided sound speed inversion in medical ultrasound. IEEE Transactions on Biomedical Engineering **67**(4), 1142–1151 (2020)
5. Hewish, A.: The diffraction of radio waves in passing through a phase-changing ionosphere. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **209**(1096), 81–96 (1951)
6. Hollman, K., Rigby, K., O'donnell, M.: Coherence factor of speckle from a multi-row probe. In: Proc. 1999 IEEE Ultrasonics Symposium (Cat. No. 99CH37027). vol. 2, pp. 1257–1260. IEEE (1999)
7. Hyun, D., Li, Y.L., Steinberg, I., Jakovljevic, M., Klap, T., Dahl, J.J.: An open source GPU-based beamformer for real-time ultrasound imaging and applications. In: 2019 IEEE International Ultrasonics Symposium (IUS). pp. 20–23. IEEE (2019)
8. ITISFoundation: Tissue properties database v4-1 (2022). https://doi.org/10.13099/VIP21000-04-01
9. Jakovljevic, M., Hsieh, S., Ali, R., Chau Loo Kung, G., Hyun, D., Dahl, J.J.: Local speed of sound estimation in tissue using pulse-echo ultrasound: Model-based approach. The Journal of the Acoustical Society of America **144**(1), 254–266 (2018)
10. Jaros, J., Rendell, A.P., Treeby, B.E.: Full-wave nonlinear ultrasound simulation on distributed clusters with applications in high-intensity focused ultrasound. The International Journal of High Performance Computing Applications **30**(2), 137–155 (2016)
11. Mallart, R., Fink, M.: Adaptive focusing in scattering media through sound-speed inhomogeneities: The van cittert zernike approach and focusing criterion. The Journal of the Acoustical Society of America **96**(6), 3721–3732 (1994)
12. Ng, G.C., Freiburger, P.D., Walker, W.F., Trahey, G.E.: A speckle target adaptive imaging technique in the presence of distributed aberrations. IEEE transactions on ultrasonics, ferroelectrics, and frequency control **44**(1), 140–151 (1997)
13. Nock, L., Trahey, G.E., Smith, S.W.: Phase aberration correction in medical ultrasound using speckle brightness as a quality factor. The Journal of the Acoustical Society of America **85**(5), 1819–1833 (1989)
14. Sanabria, S.J., Brevett, T., Ali, R., Telichko, A., Dahl, J.: Direct speed of sound reconstruction from full-synthetic aperture data with dual regularization. In: 2022 IEEE International Ultrasonics Symposium (IUS). pp. 1–4 (2022)

15. Sanabria, S.J., Ozkan, E., Rominger, M., Goksel, O.: Spatial domain reconstruction for imaging speed-of-sound with pulse-echo ultrasound: simulation and in vivo study. Physics in Medicine & Biology **63**(21), 215015 (2018)
16. Simson, W.A., Paschali, M., Sideri-Lampretsa, V., Navab, N., Dahl, J.J.: Investigating pulse-echo sound speed estimation in breast ultrasound with deep learning. arXiv preprint arXiv:2302.03064 (2023)
17. Stähli, P., Kuriakose, M., Frenz, M., Jaeger, M.: Improved forward model for quantitative pulse-echo speed-of-sound imaging. Ultrasonics **108**, 106168 (2020)
18. Vishnevskiy, V., Sanabria, S.J., Goksel, O.: Image reconstruction via variational network for real-time hand-held sound-speed imaging. In: Knoll, F., Maier, A., Rueckert, D. (eds.) Machine Learning for Medical Image Reconstruction. pp. 120–128. Springer International Publishing, Cham (2018)
19. Wild, A., Hobbs, R., Frenje, L.: Modelling complex media: an introduction to the phase-screen method. Physics of the Earth and Planetary Interiors **120**(3), 219–225 (2000)