

Enhancing sizing accuracy in ultrasound images with an alternative ADMIRE model and dynamic range considerations

Siegfried Schlunk^{a,*}, Ryan Hsi^b, Brett Byram^a

^a Department of Biomedical Engineering, Vanderbilt University, Nashville, 37232, TN, USA

^b Department of Urology, Vanderbilt University Medical Center, Nashville, 37232, TN, USA

ARTICLE INFO

Keywords:

Medical ultrasound
Beamforming
Signal processing
In vivo
Image quality
Model

ABSTRACT

Ultrasound imaging can struggle with sizing accuracy, especially when the targets have a significantly different amplitude compared to the surrounding background. In this work, we consider the challenging task of accurately sizing hyperechoic structures, and specifically kidney stones, where accurate sizing is critical for determining medical intervention. AD-Ex, an extended alternative model of our aperture domain model image reconstruction (ADMIRE) pre-processing method, is introduced and is designed to improve clutter removal and improve sizing accuracy. We compare this method against other resolution enhancing methods such as minimum variance (MV) and generalized coherence factor (GCF), and against those methods using AD-Ex as a pre-processing tool. These methods are evaluated among patients with kidney stone disease, with the task of accurately sizing the stones against the gold standard, computed tomography (CT). Stone ROI's were selected using contour maps as reference from which the lateral stone size was estimated. Among the *in vivo* kidney stone cases we processed, AD-Ex+MV had the overall lowest sizing error among the methods, with an average error of 10.8% compared to the next best method AD-Ex which had an average error of 23.4%. For reference, DAS had an average error of 82.4%. Though dynamic range was evaluated to determine optimal thresholding for sizing applications, variability between stone cases was too high for any conclusions to be drawn at this time.

1. Introduction

Sizing in ultrasound imaging can be a challenging task due to the width of the main lobe. Hypoechoic structures such as cysts and blood vessels may appear smaller and hyperechoic structures like calcifications may appear larger. A classic clinical task where this is relevant is in the diagnosis of kidney stone disease. Diagnosis of kidney stone disease is a two-part challenge for ultrasound, requiring both the correct detection and subsequently the correct sizing of a stone. Though ultrasound struggles with detection, coherence-based methods such as short-lag spatial coherence (SLSC) [1] and mid-lag spatial coherence (MLSC) [2–5] have shown promise, and color Doppler is frequently used to enhance stone “twinkling” [6,7].

In comparison, sizing has demonstrated to be the more challenging task for ultrasound beamformers, especially since the size of the stone dictates whether or not surgical intervention is necessary. It is generally accepted that stones smaller than 5 mm are able to pass naturally, while larger stones require intervention [8,9]. Ultrasound regularly overestimates the size of kidney stones by as much as 2–3 mm [10–16], which is likely to lead to an unnecessary recommendation for intervention. It has been suggested that measuring the acoustic shadow of the

kidney stone can improve accuracy [2,15,17], though the appearance of the shadow can depend on surrounding tissue and the transducer parameters [18,19], and is difficult to detect for small stones [17,20]. Other semi-automatic measurement techniques have been proposed and tested *in vitro* [3,15], but the ability to translate these methods to *in vivo* cases may be difficult due to the significantly increased complexity of surrounding tissue. As a result, the gold standard for sizing falls to computed tomography (CT) [21,22].

This work expands on our early efforts with expanded ADMIRE models in simulations [23] and will focus on the sizing task related to characterization of kidney stone disease, where the goal is to minimize the error between ultrasound measured sizes and the CT sizes, ideally attaining a margin of error for each stone that is less than or equal to $\pm 10\%$ of the CT reported size. In this work, we compare classic methods such as minimum variance (MV) and generalized coherence factor (GCF) to a new variant of ADMIRE, AD-Ex. ADMIRE, or aperture domain model image reconstruction, is a model-based beamformer that removes clutter components based on the linear physics of wave propagation, and AD-Ex aims to further improve rejection of

* Corresponding author.

E-mail address: siegfried.g.schlunk@vanderbilt.edu (S. Schlunk).

<https://doi.org/10.1016/j.ultras.2023.106952>

Received 2 September 2022; Received in revised form 10 January 2023; Accepted 7 February 2023

Available online 10 February 2023

0041-624X/© 2023 Elsevier B.V. All rights reserved.

off-axis clutter signals, which produces improved lateral performance for sizing accuracy tasks. We additionally show that this new method outperforms ADMIRE combined with MV, which we have previously demonstrated enhances lateral resolution compared to either MV or ADMIRE alone [24,25]. Since ADMIRE and AD-Ex can both be used as pre-processing steps, for example to use with MV as mentioned above, matching and improving on the performance of ADMIRE+MV enables further considerations for other beamformers to enhance other aspects of image quality. We include simulations to get a baseline understanding of our methods, but the primary focus is on the application of our methods to *in vivo* kidney stones.

We additionally consider how sizing accuracy might be related to the dynamic range of the image. Rather than strictly measure stones at traditional dynamic ranges, we show how the dynamic range of the image impacts the size of the target as measured by the beamformers presented.

2. Beamforming algorithms

All beamformers and methods were implemented in MATLAB (The MathWorks, Natick, MA, USA).

2.1. Delay-and-sum (DAS)

Delay-and-sum (DAS) is the summation of delayed channel data across M channels as given by

$$S_{\text{DAS}}(x, z) = \sum_{i=1}^M w_i(x, z) s_i(x, z). \quad (1)$$

The delayed channel data $s_i(x, z)$ corresponds to channel i , a-line index x , and discrete time index z . Apodization can be controlled by the weighting factor $w_i(z)$. Hamming apodization is often used for the weighting term to improve sidelobe performance, however, this is at the cost of broadening the main lobe, reducing spatial resolution. Since our focus for this work is on improving sizing, particularly in the case of kidney stones using ultrasound, we opt to use rectangular apodization.

2.2. Minimum variance (MV)

Minimum variance (MV) is functionally an adaptively weighted DAS image. By optimizing the weighting terms from the DAS equation to reduce off-axis interference, lateral resolution can be improved [26,27]. The optimized weights (boldface indicates vectors) are defined as

$$\mathbf{w} = \frac{\mathbf{R}^{-1} \mathbf{e}}{\mathbf{e}^H \mathbf{R}^{-1} \mathbf{e}}, \quad (2)$$

for the steering vector $\mathbf{e}$ and conjugate transpose H . The covariance matrix $\mathbf{R}$ is further defined as

$$\mathbf{R}(x, z) = \mathbf{E}[\mathbf{s}(x, z) \mathbf{s}(x, z)^H], \quad (3)$$

for the expectation $\mathbf{E}[\cdot]$ and delayed aperture signal $\mathbf{s}(x, z)$. Subarray averaging and diagonal loading methods [27] are used to get an invertible covariance matrix, with the conventional choice of subarray lengths of $L = 0.5M$ and diagonal loading equal to $\epsilon = \Delta \cdot \text{tr}(\hat{\mathbf{R}})$, where $\Delta = 1/(10L)$ [27]. The MV signal can then be estimated from the delayed channel data for each subarray $\hat{\mathbf{s}}(x, z)$ with

$$\hat{S}_{\text{MV}}(x, z) = \frac{1}{M - L + 1} \sum_{l=0}^{M-L} \mathbf{w}(x, z)^H \hat{\mathbf{s}}_l(x, z). \quad (4)$$

Since the goal of this work is improving spatial resolution to improve sizing accuracy, it should be noted that adjusting the subarray lengths and amount of diagonal loading changes the resolution of the beamformer. In particular, we generally want smaller amounts of diagonal loading, while increasing the size of the subarrays will result in improvements to resolution [27]. However, like others, we found in our previous work involving minimum variance [25] that increasing the

subarray size too much results in significant image quality degradation, which can render an image unrecognizable. As a result, we will use moderate lengths of $L = 0.5M$ and $\Delta = 1/(10L)$ for simulations, while *in vivo* we drop the subarray sizes to $L = 0.25M$. The exception to this is when we apply minimum variance after pre-processing with ADMIRE or AD-Ex (described in Section 2.6), where we found we can safely increase the subarray sizes back to $L = 0.5M$ without loss of image quality [25].

2.3. Generalized coherence factor (GCF)

Rather than weighting the channels as in MV, generalized coherence factor (GCF) weights the final DAS image. This weighting is the ratio of the energy in a chosen low-frequency region of the aperture to the total energy [28], and is calculated per pixel (x, z) with

$$\text{GCF}(x, z) = \frac{\sum_{k=0}^{M_0} |P_k(x, z)|^2}{\sum_{k=0}^M |P_k(x, z)|^2}. \quad (5)$$

The low-frequency region is defined by a cutoff index M_0 calculated from the discrete Fourier transform $P(k)$ as defined by

$$P_k(x, z) = e^{j\pi k} \sum_{i=0}^M s_i(x, z) e^{-j2\pi(ik/M)}. \quad (6)$$

By controlling the cutoff index we can range between coherence factor ($M_0 = 0$) and DAS ($M_0 = M$), and here we choose to use $M_0 = 5$ as done in previous works [29,30]. The weighted GCF image is then

$$S_{\text{GCF}}(x, z) = \text{GCF}(x, z) S_{\text{DAS}}(x, z). \quad (7)$$

2.4. Aperture domain model image reconstruction (ADMIRE)

Aperture domain model image reconstruction (ADMIRE) is a pre-processing method designed for the removal and suppression of reverberation clutter, off-axis interference, and wavefront aberration. A detailed explanation of the full algorithm is presented by Byram et al. [31] with more specifics available elsewhere [32,33], but we will reproduce as much as is necessary to understand the work here.

ADMIRE performs all processing in the frequency domain by dividing up broadband signals into narrowband components for processing. This is done by subdividing the delayed channel data axially into many overlapping windows, and applying a short-time Fourier transform (STFT). We generally use 90% overlap between windows to maintain good speckle texture. This produces an aperture signal for each a-line and primary frequency component of the windowed broadband signal. Each aperture domain signal is analyzed against a physics-based model that contains the predicted aperture domain signal responses for scatterers throughout the imaging space. These responses can be calculated due to the linear nature of sound wave propagation, and in theory any received signal can be represented as the linear combination of these responses with the equation

$$p_s(x; t, \omega) = \sum_{n=0}^{N-1} A(x; x_n, z_n, \tau_n, \omega) e^{jk\tau(x; x_n, z_n, \tau_n)}. \quad (8)$$

Here, x is indexing the aperture location and ω defines the post-STFT frequency of the signal. For the signal arriving at the aperture at time t , there are a total of N responses from scatterers in the medium arriving at that time. The wavefront delay $\tau(x; x_n, z_n, \tau_n)$ is for a signal originating from (x_n, z_n) at time τ_n , and k is the wavefront number. Finally, a measure of amplitude modulation across the aperture $A(x; x_n, z_n, \tau_n, \omega)$ is included, which is based on effects of the STFT windows and element sensitivity.

By individually estimating these scatterer responses, we can combine all of them into a model matrix, $\mathbf{X}$, which can be used to represent any frequency-domain aperture domain signal, y , with the equation

$$y = \mathbf{X}\beta, \quad (9)$$

Table 1
ADMIRE and AD-Ex Model space parameters.

Parameter	Value
α	0.9
λ	(0.00189) y_{RMS}
X_{clutter} Sampling	$3.577 \text{res}_{\text{lateral}} \cdot 7.154 \text{res}_{\text{axial}}$
X_{clutter} Size	Full Imaging Space
X_{ROI} Sampling	$0.179 \text{res}_{\text{lateral}} \cdot 0.715 \text{res}_{\text{axial}}$
X_{ROI} Ellipse Radii	$(0.5 \text{res}_{\text{lateral}}, \text{res}_{\text{axial}})$
X_{extended} Sampling	$0.358 \text{res}_{\text{lateral}} \cdot 0.715 \text{res}_{\text{axial}}$
X_{extended} Ellipse Radii	$(\text{res}_{\text{lateral}} \cdot 2 \text{res}_{\text{axial}})$

where every received signal y corresponds to some location (x_n, z_n) and frequency ω , and we are solving for β , the unknown set of model coefficients that will reveal the origins of the scatterer responses that combine to make the received signal. Fig. 1A shows a visual example of how signals in the model $X = [X_{\text{clutter}} X_{\text{ROI}}]$ could correspond to specific physical locations in the field of view. By design, the matrix consists of both clutter signals X_{clutter} , which includes image degradation sources such as reverberation clutter and off-axis interference, and region of interest (ROI) signals X_{ROI} , which is all the signals originating near the target (x_n, z_n) . The model coefficients β then represent the specific signals that linearly combine to form y . By solving for β and zeroing those coefficients corresponding to signals in X_{clutter} , the decluttered aperture signal can be reconstructed with

$$y_{\text{decluttered}} = X_{\text{ROI}} \beta \quad (10)$$

We can specify which signals we want to keep by how we define the ROI, an ellipse, around the target (x_n, z_n) . The major and minor radii of this ellipse are calculated as $c_{\text{axial}} \text{res}_{\text{axial}}$ and $c_{\text{lateral}} \text{res}_{\text{lateral}}$, with c_{axial} and c_{lateral} scalars based on how large we desire the ROI to be in the axial and lateral dimensions. The axial sampling is approximated with $\text{res}_{\text{axial}} \approx 2 \text{res}_{\text{lateral}}$, and the lateral sampling is estimated as $\text{res}_{\text{lateral}} \approx \lambda_w z_n F\{|p_s(x_n, z_n, 0)|\}_{BW}$, where λ_w is the wavelength and $F\{|\cdot|\}_{BW}$ is the lateral bandwidth at (x_n, z_n) . The specific choices for this application are included in Table 1, with the scalars chosen empirically from previous testing. As defined here, the ROI for the model is determined based on physical ultrasound parameters, and is completely independent of the actual content in the ultrasound image.

However, the inverse matrix calculation that would be used to solve (9) is ill-posed due to the much smaller size of the aperture signal y compared to the much larger size of the model X . Instead, the elastic-net regularization technique [34] is used to solve for the model coefficients. The optimization equation is

$$\hat{\beta} = \arg \min_{\beta} (\|y - X\beta\|^2 + \lambda(\alpha\|\beta\|_1 + (1 - \alpha)\|\beta\|_2^2/2)), \quad (11)$$

where $\|\beta\|_1$ and $\|\beta\|_2$ are the L1 and L2 norm, and α can be set between 0 and 1 to control the weighting between them. The degrees of freedom of the solution can be controlled by a separate regularization parameter, λ [35]. The chosen values for these parameters are included in Table 1, based on values from previous work [31,32].

With the model coefficients estimated by the elastic-net, the decluttered aperture domain signal can be calculated with (10) as described. The decluttered signal can then be returned to the time domain via the inverse STFT [36]. Since this entire process is performed along the channel dimension, the result is decluttered channel data, where post-processing methods such as DAS, MV, and GCF can still be applied. Computation using these models required the use of the Vanderbilt University ACCRE computing center, where *in vivo* cases were generally split among 100 processors, each taking between 30 to 90 min to compute, varying based on the size of the model which increases with increased depth. Though real-time implementations of ADMIRE have been created [37,38], the model creation process in that version is different and is currently incompatible with this work.

2.5. ADMIRE-extended (AD-Ex)

AD-Ex is an alternative model setup for ADMIRE originally implemented with iterative ADMIRE [23,30,39]. Due to the interaction of the elastic-net regularization and our model, we occasionally run into the problem of certain signals not being fit well. Specifically, since the elastic-net balances between L1 and L2 regularization, we have some of the benefits and detriments of both. L1 pushes towards sparse, selective solutions, while L2 pushes to include all predictors when there is correlation among predictors. Predictors in close proximity can be highly correlated, and thus the use of L2 in the regularization results in many of these predictors being selected together, which can be an issue if they are located near the border of the ROI. The highly-sampled ROI in the model is attractive to the elastic-net for fitting, which can result in somewhat correlated non-ROI signals being fit with ROI predictors, and therefore not being correctly removed. This is especially problematic for off-axis clutter signals originating from near, but not in, the ROI.

The model solution we chose is $X = [X_{\text{clutter}} X_{\text{ROI}} X_{\text{extended}}]$, shown in Fig. 1B. We designed a small extended clutter region surrounding the ROI that has a sampling rate that falls in between that of the ROI and the rest of the clutter region. The specific sampling and size of the extended region is included in Table 1. We consider both X_{clutter} and X_{extended} as clutter regions to be removed. This results in clutter signals that were being fit in the ROI being fit into this extended clutter region instead, resulting in the correct removal of these signal components. In testing, the extended model did not have a significant impact on computation time, as the additional model predictors are negligible compared to the size of the base ADMIRE model.

In summary, for clutter signals that originate from inside the extended region, those predictors are more likely to be correlated with predictors in the ROI due to close physical proximity. This means that normal ADMIRE is unlikely to fully reject those signals, meaning the acceptance region is effectively larger than we have actually designed it. However, using the extended region will help to fit these clutter signals more accurately, and critically to reduce the likelihood of the elastic-net preferring ROI predictors. This is shown in the simulation included in Fig. 1, where the off-axis scatter marked by the 'x' is partially fit using the ROI acceptance region when using the normal ADMIRE model, which would result in incomplete rejection. However, using the extended model does prevent the ROI from being used to fit the clutter signal. This will improve rejection of off-axis clutter signals and should improve lateral resolution and performance when using the extended model. By using lower sampling compared to the actual ROI, this should prevent the reverse problem of ROI signals being fit into the extended region.

2.6. Post-ADMIRE processing

Due to the nature of the reconstruction step with ADMIRE, the original dimensionality of the input data can always be recovered after denoising. This makes ADMIRE a powerful option as a pre-processing tool that can be used before other beamformers. We have shown in previous work that MV can benefit from pre-processing with ADMIRE for improving lateral resolution [24,25,40] due to the denoising that ADMIRE provides, so we include here MV applied after pre-processing with ADMIRE (AD+MV) to compare against AD-Ex. Since AD-Ex is a modification of ADMIRE, it similarly can function as a pre-processing step, so we also include both MV and GCF applied after pre-processing with AD-Ex (AD-EX+MV and AD-EX+GCF) to see how they compare against AD-Ex alone for completeness.

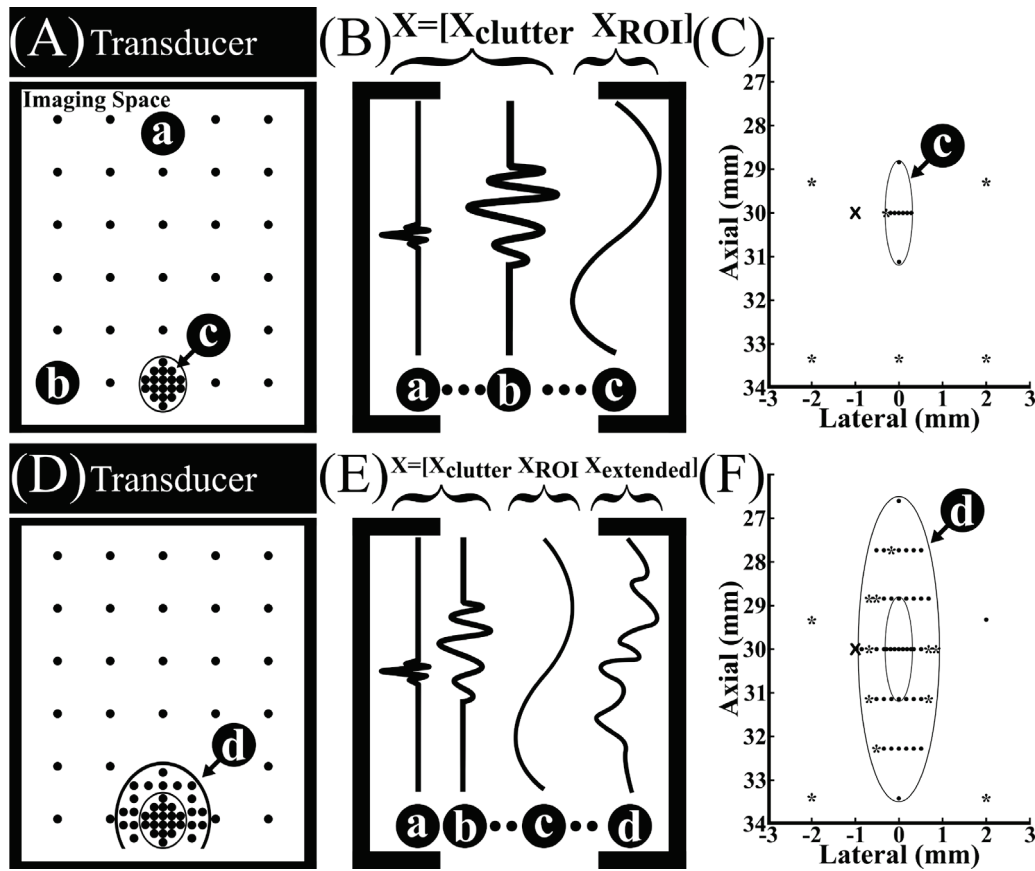


Fig. 1. (A-B) An example ADMIRE model X , which consists of signals from clutter locations (sparsely sampled, e.g. 'a' and 'b'), and signals from region of interest locations (highly sampled, e.g. 'c'). By solving the matrix equation, we can estimate which physical sources in the modeled imaging space make up the received echo. Signals 'a' and 'b' represent reverberation clutter signals and off-axis signals, respectively, while 'c' is an example of a signal from the ROI. (D-E) A modified model, called here the AD-Ex model, includes an extension that is an intermediate clutter region $X_{extended}$ composed of signals represented by 'd'. The proposed intermediate clutter region is sampled at half that of the ROI, but still more than the remaining clutter. This helps to ensure proper classification and removal of signals originating near, but not in, the ROI. (C, F) Plots that show a simulation example where a single off-axis scatterer (marked by the 'x') is simulated outside of the ROI (shown as the small circled area). The dots indicate physical locations where signals are modeled, and the asterisks show locations that are being used by the elastic-net to fit the observed signal. In (C), the normal model from (A) is used, and some ROI predictors are used by the elastic-net in order to best-fit the observed signal. However, in (F) when the extended model from (D) is used, the same signal is fit only using clutter predictors from the background and extended regions, resulting in more complete removal of the off-axis signal.

3. Methods

3.1. Simulated phantoms

We used Field II [41,42] to simulate $n=6$ 5 mm hyperechoic cysts at amplitudes of 40, 50, 60, and 70 dB as rudimentary equivalents to kidney stones at our targeted size of interest. We additionally simulated single point targets with 60 dB signal-to-noise ratio (SNR) of added Johnson–Nyquist noise to perform resolution comparisons between methods. The simulation parameters for both simulations are detailed in Table 2.

We estimated the lateral amplitude profiles of the simulated cysts by averaging the amplitude radially within $\pm 10^\circ$ of the lateral axis, as is shown in Fig. 2. For each method and cyst we calculated the lateral cyst size from the profiles at the 30 dB and 6 dB thresholds. The 30 dB threshold represents how a sonographer would normally size from a b-mode image, since 30 dB or lower thresholds all measure equivalently, as can be seen from the amplitude plot in the figure. The 6 dB threshold is an alternative based on how the full width at half maximum (FWHM) is measured for resolution metrics. For the simulated point targets, the lateral point-spread functions were evaluated for each method.

3.2. In vivo kidney stone data

We captured *in vivo* kidney data from patients with one or more kidney stones using a Verasonics Vantage Ultrasound System (Verasonics,

Table 2

Field II simulation parameters for contrast target phantoms.

Parameter	Value
Number of elements	117
Number of mathematical elements (lateral)	7
Number of mathematical elements (elevation)	11
Element height	4 mm
Element width	0.254 mm
Kerf	0.003 mm
Lateral pitch	0.257 mm
Center frequency (f_c)	3 MHz
Sampling frequency (simulation) (f_s)	640 MHz
Sampling frequency (downsampled) (f_s)	40 MHz
Bandwidth	60%
Transmit focal depth	3 cm
Transmit/receive f-number	1

Inc., Kirkland, WA) with a C5-2 curvilinear transducer. A curvilinear plane wave synthetic focusing acquisition [43] was used, which allows for focusing at all depths using synthetic transmit focuses that are placed throughout the image. A center frequency of 4.1667 MHz was used to acquire 64 angles uniformly spaced spanning 37° . For every patient and stone, a previous CT scan was available from which the stone size was measured, and the longest axis was determined. This allowed the physician to acquire the ultrasound data such that the longest axis was in the lateral dimension, which is the dimension used

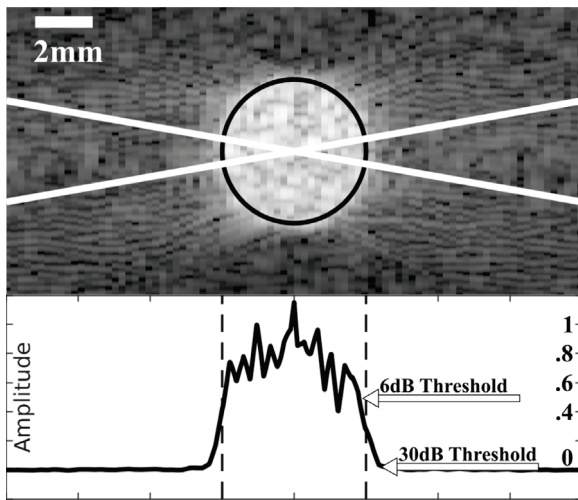


Fig. 2. (Top) DAS of an example 40 dB simulated 5 mm cyst, with the solid black line indicating the true cyst boundary. The cyst boundary was estimated from the radial region indicated by the white lines. (Bottom) plot of radial average amplitude, with the dashed lines indicate the true edges of the cyst. We indicate two thresholds, 6 dB and 30 dB, to show how they would underestimate and overestimate cyst size, respectively.

clinically to measure stone size. The axial axis is generally avoided for sizing since it can be impacted by reverberation artifacts such as comet trails. We acquired $n=8$ kidney stone images across 7 different patient volunteers that had recent CT images on file to use as the gold standard. In total, CT stone sizes ranged from 2.3 to 12.2 mm. The average BMI was 30.6 ± 7.1 kg/m², with four of the stones originating from patients being categorized as obese (≥ 30), two as overweight (≥ 25 and <30), and two as being within the normal range (≥ 18.5 and <25). The Vanderbilt University IRB approved the protocols used to acquire the data (IRB# 170001).

3.3. Kidney stone image metrics

First, for every stone and method, we needed to select the stone in the image for sizing and image quality purposes. We first had a physician manually size 30 dB dynamic range DAS images of each stone. Then a contour map for each DAS case and the physician measurement was used to manually draw an ROI of the stones in those DAS cases. An ROI was then manually drawn for each other method using the respective contour map as a guide, but without a physician measurement as reference. Since this work focuses solely on the sizing task of diagnosis, all stone locations were already known when measuring the images, but the stone size was unknown. From Fig. 2, sizing at 30 dB or a more traditional 50 dB should be comparable, so we chose to use 30 dB for all sizing expecting that it would aid in stone visibility. Note that this stone ROI is unrelated to the ROI in the ADMIRE algorithm, and they are chosen completely independent of each other.

The stone size was automatically estimated from each stone ROI as the longest lateral distance. The lateral axis is chosen even in cases where the axial dimension appears larger as it is consistent with clinical procedures, and attempts were made when acquiring data to make the longest stone axis be along this dimension. The background region is also automatically selected as a circular region centered on the middle of the stone ROI, with a total area of 2.5 times the area of the stone ROI (to compensate for the non-standard shapes of the kidney stones). This process is challenging due to the extremely wide variety of shapes kidney stones can take, as shown in Fig. 3, which can make it difficult to differentiate an oddly-shaped large kidney stone versus a smaller stone with adjacent bright tissue. As a result, we tried to balance the physician measurements with the contour maps to be as objective as

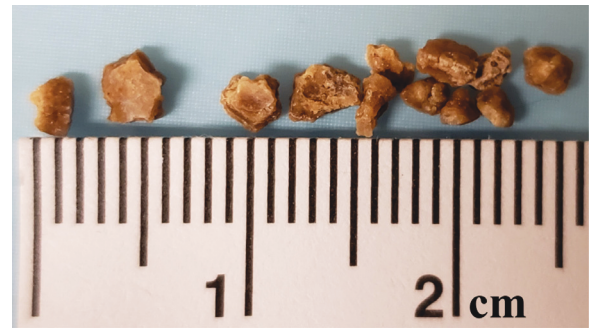


Fig. 3. Examples of small ex vivo kidney stones, to demonstrate the wide variety of shapes they can take.

possible with the ROI selections, though we have also included physician measurements for AD+MV, AD-Ex, AD-Ex+MV, and AD-Ex+GCF to compare against these manually drawn ROI measurements. All of the stones and the ROI selections can be found in the supplementary materials for transparency. An example of this ROI selection and sizing process is shown in the top row of Fig. 5.

ROI-based measurements and physician measurements were compared against the provided CT measurements and both measured size and percent error compared to CT were reported. For completeness, the contrast ratio and generalized contrast-to-noise ratio (gCNR) [44,45] were calculated from the stone ROI selected and a background region chosen. Contrast ratio was defined as

$$\text{contrast ratio} = -20 \log_{10} \left(\frac{\mu_{\text{ROI}}}{\mu_{\text{background}}} \right) \quad (12)$$

where μ is the mean value of the indicated region of the enveloped, uncompressed data. Overlap for gCNR was estimated using the standard 100 histogram bins.

To evaluate how the dynamic range of the b-mode image may impact sizing accuracy, we estimated the stone ROI for each case at 1 dB increments on dynamic ranges ranging from 30 to 6 dB. This allowed us to plot the measured stone size versus the dynamic range threshold at which it was measured, giving us an estimate of what dynamic range could be used for each case to accurately measure that specific stone.

4. Results

4.1. Simulated cysts and point targets

The average estimated widths of the 5 mm simulated hyperechoic cysts are included in Table 3. The 30 dB threshold measurements show that all methods tend to overestimate the size of the cyst by 0.5–1.5 mm, more than 10%. In comparison, at 6 dB all methods underestimated the cyst size, implying that if we measured at some intermediate threshold we would size cysts accurately. However, each method is off by varying amounts, which means that the accurate threshold would be different for each method. For example, DAS and ADMIRE are fairly accurate at 6 dB, but MV underestimates significantly. The measurements are mostly consistent for different cyst amplitudes, with a minor trend of larger estimates for brighter cysts.

The point spread functions for the simulated single point targets are included in Fig. 4. The MV, AD+MV, and AD-Ex+MV methods are noticeably more narrow at the apex of the PSF compared to other methods. AD-Ex separately shows significant improvement by lowering the amplitude of the side lobes, which is maintained when used in combination with MV and GCF. AD-Ex+MV shows additional main lobe improvement compared to MV, AD-Ex, or AD+MV at all amplitudes.

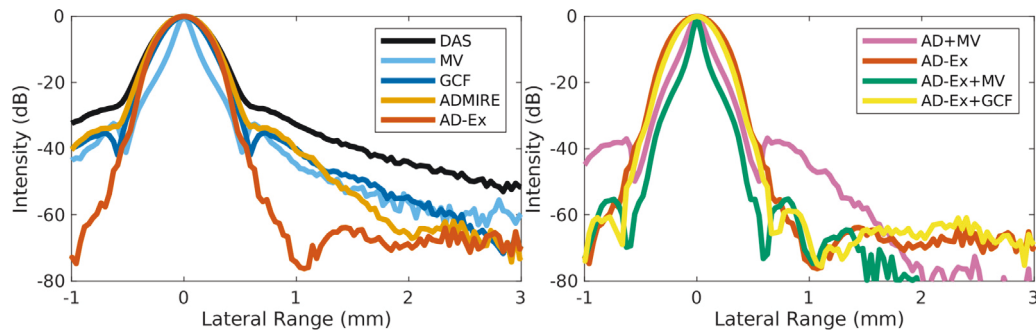


Fig. 4. Point spread functions of simulations of a single point target with 60 dB SNR of added thermal noise. DAS, GCF, and ADMIRE all have similar profiles, while MV has a narrower profile in the range of -20 to 0 dB. AD-Ex has a similar profile to ADMIRE, but narrows at levels below -40 dB. AD+MV and AD-Ex+MV show a narrower profile compared to either method alone at all levels, while AD-Ex+GCF is nearly identical to AD-Ex.

Table 3

Measured sizing error of 5mm cysts at various amplitudes (Measured - true).

Measured size error at 30 dB threshold (mm)				
Cyst	40 dB	50 dB	60 dB	70 dB
DAS	1.23 ± 0.22	1.29 ± 0.22	1.29 ± 0.23	1.43 ± 0.25
MV	0.86 ± 0.08	0.90 ± 0.06	0.92 ± 0.07	0.92 ± 0.08
GCF	0.83 ± 0.07	0.83 ± 0.07	0.83 ± 0.07	0.89 ± 0.14
ADMIRE	0.87 ± 0.06	0.90 ± 0.06	0.90 ± 0.08	0.98 ± 0.17
AD+MV	0.74 ± 0.08	0.78 ± 0.06	0.79 ± 0.06	0.83 ± 0.09
AD-Ex	0.74 ± 0.09	0.76 ± 0.07	0.77 ± 0.06	0.81 ± 0.10
AD-Ex+MV	0.56 ± 0.11	0.58 ± 0.10	0.59 ± 0.10	0.61 ± 0.12
AD-Ex+GCF	0.69 ± 0.12	0.71 ± 0.11	0.71 ± 0.11	0.75 ± 0.13
Measured size error at 6 dB threshold (mm)				
Cyst	40 dB	50 dB	60 dB	70 dB
DAS	-0.23 ± 0.34	-0.22 ± 0.33	-0.22 ± 0.33	-0.28 ± 0.38
MV	-1.05 ± 0.43	-1.01 ± 0.50	-1.01 ± 0.50	-1.01 ± 0.50
GCF	-0.73 ± 0.67	-0.73 ± 0.67	-0.73 ± 0.66	-0.77 ± 0.70
ADMIRE	-0.23 ± 0.35	-0.22 ± 0.34	-0.22 ± 0.34	-0.28 ± 0.39
AD+MV	-0.83 ± 0.64	-0.82 ± 0.64	-0.82 ± 0.64	-0.83 ± 0.65
AD-Ex	-0.31 ± 0.58	-0.31 ± 0.58	-0.31 ± 0.58	-0.36 ± 0.60
AD-Ex+MV	-0.85 ± 0.64	-0.75 ± 0.68	-0.75 ± 0.68	-0.78 ± 0.71
AD-Ex+GCF	-0.47 ± 0.64	-0.47 ± 0.64	-0.47 ± 0.64	-0.66 ± 0.73

4.2. In vivo kidney stone results

4.2.1. ROI-based results

Fig. 5 shows an example of the *in vivo* sizing process for DAS (top row) and the resulting region selections for the other methods (bottom row) for kidney stone Case 1. The red line with arrows indicates the physician estimated size from the 30 dB dynamic range b-mode image, while the blue line with tick marks indicates the ROI estimated size of the stone. The remaining b-mode and region selections for all other stone cases can be found in the supplementary materials, including physician and ROI size estimates. Note that only DAS, AD+MV, AD-Ex, AD-Ex+MV, and AD-Ex+GCF were manually sized by the physician as the other methods did not produce sufficient improvements to be considered.

Table 4 shows the measured stone size and percent error compared to CT when using the manually drawn ROI, while the physician measurements are included in Table 6 and will be discussed later in the section. The BMI of the patient, depth of the stone, and CT measurements are included in each table for quick reference. Table 5 includes the contrast ratio and gCNR based on the ROI of each stone. Stone cases are listed in ascending order of CT measured size (i.e. Case 1 is the smallest stone and Case 8 is the largest).

The ROI-based measurements in Table 4 show that the stones were very rarely underestimated, and those instances were by less than 5%. On average, DAS reported 82.4% error, while the combined ADMIRE methods all reported <30%, an improvement compared to the other methods. The combination with MV improved accuracy and lowered

variance for both ADMIRE and AD-Ex, with AD-Ex+MV reporting the lowest average error at 10.8% and the second lowest variance among the methods, only narrowly being beaten by AD+MV. The improvement of combining MV with either variant of ADMIRE is particularly more effective than when MV is used by itself, which only provides a minor improvement in accuracy and a small improvement in variance compared to DAS. GCF fairs a bit differently, performing better alone (producing improvements compared to DAS and MV) than in combination with AD-Ex (equivalent performance in most cases, with worse performance in Case 4). Fig. 6 shows a summary of the ROI measured sizes, though DAS and MV are not included to improve readability of the figure. Overall, AD-Ex+MV sizes five of the eight cases within an error of $\pm 10\%$ and is the most accurate method in two of the remaining three cases. This is followed by AD-Ex which manages an error of less than $\pm 10\%$ in four cases, and AD-Ex+GCF at three cases.

The contrast ratio and gCNR calculated from the stone ROI for each case and method are included in Table 5, giving us a representation of the visibility of these stones, with higher values suggesting easier visibility. Generally, GCF or AD-Ex+GCF had the greatest contrast ratio and gCNR of all the methods, and both report consistently higher than DAS and MV. AD-Ex reported higher contrast ratio and gCNR compared to AD+MV in all but one case (Case 3), and AD-Ex+MV and AD-Ex+GCF performed similarly to AD-Ex alone, though sometimes producing minor improvements. This means that the combination of MV or GCF with AD-Ex appears to maintain the image quality improvements that AD-Ex is producing.

Fig. 7 shows an attempt to optimize the dynamic range threshold for measuring stones for DAS, GCF, and AD-Ex. The numbered triangles indicate the stone case and the threshold at which the stone is measured accurately, i.e. where the solid ultrasound measurement line meets the dashed CT size line. DAS appears to be accurate mostly in the range of -6 to -12 dB, while GCF and AD-Ex are much more varied. These plots generally agree with our cyst simulations, showing more overestimation towards -30 dB, and less or even underestimation towards -6 dB. It also in agreement with our decision to make measurements at -30 dB, since the curves flatten out considerably in that range for all methods and stones, meaning the sizing would be fairly similar to a more normal b-mode image at 50 or 60 dB. However, the large variability of the optimal thresholds, especially between methods, makes it difficult to determine if there exists an optimal threshold for stone sizing. As a result, choosing the 30 dB threshold for the rest of the measurements was the only objective way we could compare each method.

Considering both sizing accuracy and contrast ratio and gCNR, the AD-Ex-based methods, AD+MV, and GCF reported better accuracy and visibility compared to the other methods, with AD-Ex+MV on top due to better sizing accuracy. Since MV primarily narrows the main lobe between -20 and 0 dB, it did not provide a noticeable improvement compared to DAS when measuring on a 30 dB dynamic range, and when measuring at -6 dB as we did for the simulated cysts, DAS was already

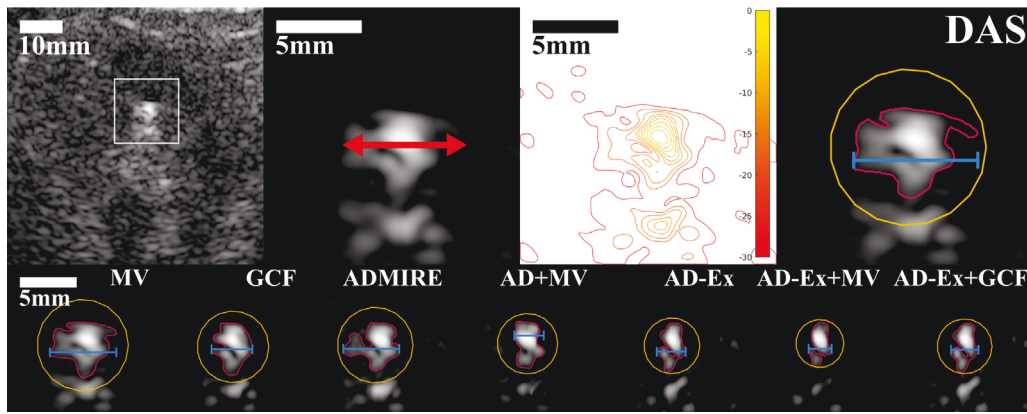


Fig. 5. Example of the sizing process for *in vivo* kidney stone Case 1 (CT size of 2.3 mm). The top row shows the process for DAS, starting with the full kidney image on a 50 dB scale. The white square in the full image indicates the region that is shown in the following zoomed images. These zoomed images are shown at 30 dB and were used for sizing. The red arrows show the physician-marked stone size, which was used in combination with the contour map to manually circle the stone region for measurements, marked as the red outline. Ultrasound when used for sizing uses the lateral axis for accuracy, so the stone size is the longest lateral extent of the ROI, marked as the blue line. The yellow line shows the region used for the background, which is automatically selected as a circular region double the area of the stone region. The bottom row shows the end results for the same case for MV, GCF, ADMIRE, AD+MV, AD-Ex, AD-Ex+MV, and AD-Ex+GCF.

Table 4
Patient statistics and ROI-based size measurements for *in vivo* kidney stones.

Patient statistics									
Case	1	2	3	4	5	6	7	8	
BMI (kg/m ²)	24.4	22.6	25.5	35.3	36.9	42.3	32.2	25.5	
Depth (cm)	4.4	4.3	5.3	8.4	13.5	8.6	12.4	7.3	
CT (mm)	2.3	3.4	4.7	5.4	5.9	6.3	7.1	12.2	
ROI measured size (mm)									
Case	1	2	3	4	5	6	7	8	
DAS	7.3	5.4	10.1	6.2	12.0	8.5	14.6	14.1	
MV	6.8	5.6	10.0	6.4	12.1	8.8	14.5	14.0	
GCF	4.1	4.5	9.0	5.7	9.3	6.7	12.2	12.4	
ADMIRE	5.7	5.5	9.0	6.8	8.5	7.4	14.1	12.6	
AD+MV	3.0	4.0	7.5	6.9	7.7	7.2	10.0	12.7	
AD-Ex	3.0	3.3	7.7	5.5	8.8	6.3	11.0	11.7	
AD-Ex+MV	2.3	3.4	6.7	6.6	5.6	6.0	9.7	11.8	
AD-Ex+GCF	2.8	3.3	7.5	7.9	8.9	6.0	11.0	11.7	
Percent error compared to CT (%)									
Case	1	2	3	4	5	6	7	8	Average
DAS	215.2	58.8	114.9	13.9	102.5	34.1	104.9	15.2	82.4 ± 67.3
MV	193.5	63.2	112.8	17.6	104.2	39.7	104.2	14.3	81.2 ± 60.0
GCF	80.4	32.4	90.4	4.6	57.6	6.3	71.8	1.6	43.2 ± 36.5
ADMIRE	147.8	60.3	90.4	26.9	43.2	17.5	97.9	3.3	60.9 ± 48.5
AD+MV	30.4	17.6	59.6	27.8	29.7	13.5	40.1	3.7	27.8 ± 17.2
AD-Ex	28.3	-2.9	62.8	0.9	49.2	-0.8	54.2	-4.1	23.4 ± 28.6
AD-Ex+MV	2.2	-1.5	42.6	21.3	-5.9	-4.8	36.6	-3.7	10.8 ± 19.8
AD-Ex+GCF	21.7	-4.4	59.6	46.3	50.0	-4.8	54.2	-4.1	27.3 ± 28.5

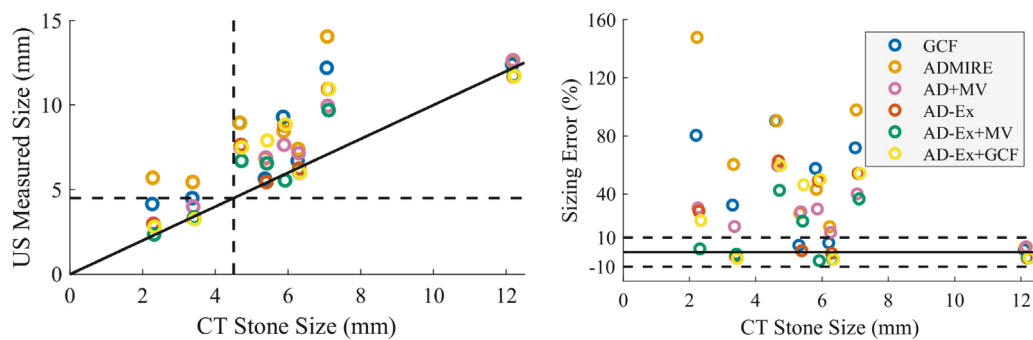


Fig. 6. (Left) US measured stone size versus CT stone size for all cases. Note that the methods are slightly staggered for visibility. The dashed lines are positioned at stone sizes of 4.5 mm, which round up clinically to 5 mm. (Right) percent sizing error versus CT stone size for all cases. The dashed lines are positioned at ±10%.

underestimating true cyst size which makes MV underestimate by an even greater amount. Since clinical measurements would be made more

in the range of 30 dB or higher, it is reasonable to expect that MV would not demonstrate an improvement for sizing *in vivo*. However,

Table 5
Contrast ratio and gCNR for *in vivo* kidney stones.

Stone contrast (dB)								
Case	1	2	3	4	5	6	7	8
DAS	15.51	11.35	14.92	7.52	13.03	14.83	11.66	12.54
MV	14.84	10.50	14.82	5.77	12.85	14.05	11.54	12.40
GCF	24.04	23.29	23.33	13.92	19.67	20.07	21.54	19.13
ADMIRE	19.71	14.23	18.32	5.85	14.99	18.44	15.20	15.16
AD+MV	23.61	17.37	19.54	6.10	14.01	17.12	13.62	14.83
AD-Ex	25.34	20.72	15.62	11.61	20.43	19.85	22.42	18.04
AD-Ex+MV	27.29	22.50	18.23	11.85	20.33	20.61	21.84	17.25
AD-Ex+GCF	26.34	22.21	15.71	11.15	20.84	20.84	23.90	18.41
Stone gCNR								
Case	1	2	3	4	5	6	7	8
DAS	0.78	0.72	0.77	0.59	0.65	0.73	0.54	0.70
MV	0.78	0.66	0.79	0.48	0.64	0.70	0.54	0.73
GCF	0.91	0.91	0.85	0.63	0.80	0.85	0.85	0.81
ADMIRE	0.86	0.76	0.84	0.44	0.75	0.82	0.72	0.76
AD+MV	0.90	0.74	0.86	0.38	0.19	0.81	0.04	0.64
AD-Ex	0.94	0.88	0.69	0.57	0.84	0.84	0.82	0.78
AD-Ex+MV	0.94	0.81	0.75	0.57	0.90	0.84	0.83	0.70
AD-Ex+GCF	0.94	0.88	0.69	0.54	0.84	0.87	0.87	0.77

when combined with ADMIRE or AD-Ex, the simulations showed that the main lobe is further narrowed over a much larger range, which does result in improvements *in vivo*.

4.2.2. Physician-based results

The physician measurements are included in Table 6, and the comparison of those values to the ROI-based measurements are included in Table 7. Measurements were not made for MV, GCF, and ADMIRE alone as they were not significantly better compared to DAS, or as good as the included methods. Since the initial DAS ROI's were measured with reference to the physician measurements, the DAS ROI-based and physician measurements are very similar. However, in all other cases the physician measurements are consistently smaller compared to the ROI-based measurements. This results in AD+MV measuring more accurately compared to the ROI's, and for many of the AD-Ex methods actually underestimating the CT stone size. This results in an increased variance in percent error compared to the ROI measurements. Based on these results, AD+MV measures within $\pm 10\%$ error in four of eight cases, followed by AD-Ex and AD-Ex+GCF with only two of eight cases.

We also include Fig. 8, which shows a visual example of Cases 1, 5, and 8 to compare how the measurements differ for DAS, AD+MV, and AD-Ex+MV. Case 8 is perhaps the largest difference between the two methods, where the physician selected an entirely different part of the structure for AD-Ex based methods, resulting in a significantly smaller estimated stone size, and a large underestimation of that size compared to CT. With that exception, the differences in the other cases are primarily due to how much of the edges of the stone are included, especially those parts at lower amplitudes that could be confused with background tissue. In the DAS case, the physician tended to include more of the peripheral structures (and potentially tissue) if the structures appeared to be connected. However, as the background was increasingly suppressed by ADMIRE and AD-Ex, the opposite became true where they tended to not include those peripheral structures, while the ROI-based method using the contour maps did. This results in consistently smaller stone measurements for these methods compared to using the ROI, resulting in an overall and consistent shift in the absolute error.

5. Discussion

When comparing the resolution of different beamformers, the FWHM is a common performance metric, but for clinical sizing applications where the dynamic range is much larger, it may not be as

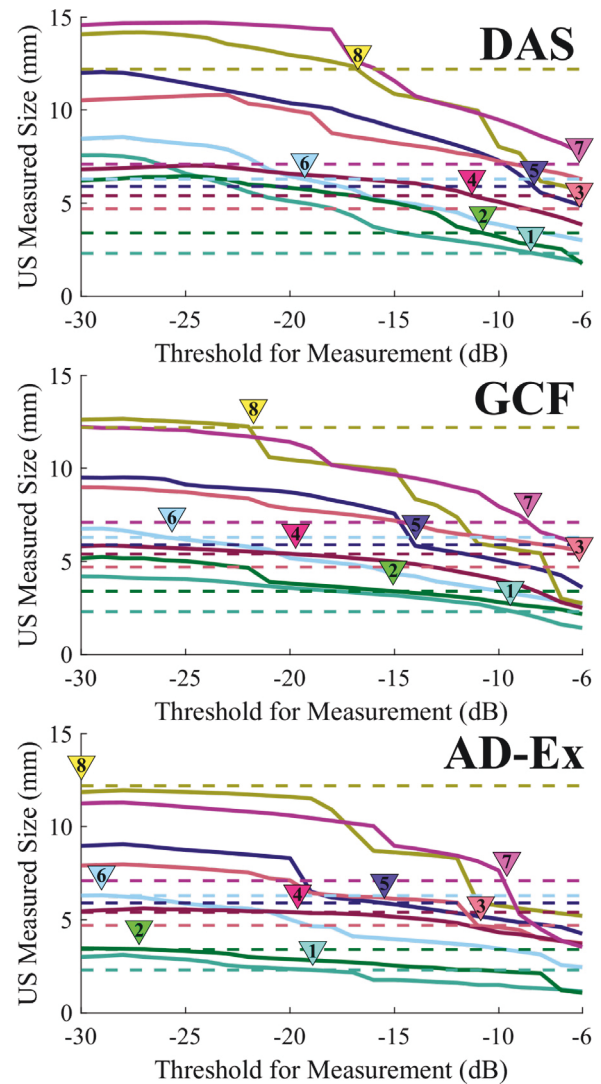


Fig. 7. Examples of measured stone size for DAS, GCF, and AD-Ex for each stone at thresholds from -30 to -6 dB. The solid lines indicate the measured size at each threshold, and the dashed lines indicate the CT measured stone size. The colored arrows indicate the location where the solid and dashed lines intersect, i.e. the threshold where the US method measurement is equal to CT stone size.

useful to gauge performance. For example, ADMIRE's normal ROI is not designed to enhance the FWHM compared to DAS, but by reducing clutter and improving contrast ratio it generally performs better at sizing compared to even MV. To further improve ADMIRE's accuracy, we knew that addressing the off-axis interference caused by mis-fitting nearby signals was important. An obvious consideration was to only use L1 for regularization since L2 is a driving factor in this fitting problem, but this has significant drawbacks of degrading image quality. We also considered other solutions, such as increasing the sampling of the clutter region or adjusting the size of the ROI to compensate, but this often resulted in dramatically increased computation time or increased variance in the image or other unexpected or undesired image quality changes. AD-Ex demonstrated the best balance, and we chose the smallest region possible for the extended region that did not result in large increases to image variance while still successfully rejecting off-axis clutter correctly. Generally, increasing the size of the extended region did not further improve sizing accuracy in our testing, and in some cases it worsened the variance of the results.

The point target simulations suggest that, in ideal conditions, MV improves the FWHM while AD-Ex improves main lobe performance at

Table 6Patient statistics and Physician size measurements for *in vivo* kidney stones.

Patient statistics									
Case	1	2	3	4	5	6	7	8	
BMI (kg/m ²)	24.4	22.6	25.5	35.3	36.9	42.3	32.2	25.5	
Depth (cm)	4.4	4.3	5.3	8.4	13.5	8.6	12.4	7.3	
CT (mm)	2.3	3.4	4.7	5.4	5.9	6.3	7.1	12.2	
Physician measured size (mm)									
Case	1	2	3	4	5	6	7	8	
DAS	7.1	5.5	10.2	6.1	11.2	8.5	15.9	14.0	
AD+MV	2.5	4.2	7.5	5.4	5.4	4.7	9.2	12.1	
AD-Ex	2.1	2.5	6.6	5.8	8.4	5.1	10.4	6.9	
AD-Ex+MV	1.7	2.6	6.6	6.4	4.7	4.1	9.3	7.2	
AD-Ex+GCF	2.1	2.5	6.6	5.8	8.4	4.5	10.7	7.0	
Percent error compared to CT (%)									
Case	1	2	3	4	5	6	7	8	Average
DAS	208.4	62.4	116.2	12.6	89.8	34.4	124.6	14.8	82.9 ± 66.6
AD+MV	9.6	24.7	60.0	0.7	-7.8	-24.9	29.4	-1.0	11.3 ± 26.2
AD-Ex	-6.7	-27.1	41.3	6.7	42.8	-18.5	46.5	-43.6	5.2 ± 34.9
AD-Ex+MV	-27.0	-24.7	41.3	18.5	-20.5	-34.4	30.6	-41.0	-7.1 ± 32.1
AD-Ex+GCF	-6.7	-27.1	41.3	6.7	42.8	-29.1	50.1	-43.0	4.4 ± 36.7

Table 7Comparison of sizing estimates of ROI method versus Physician of *in vivo* kidney stones.

Patient statistics									
Case	1	2	3	4	5	6	7	8	
BMI (kg/m ²)	24.4	22.6	25.5	35.3	36.9	42.3	32.2	25.5	
Depth (cm)	4.4	4.3	5.3	8.4	13.5	8.6	12.4	7.3	
CT	2.3	3.4	4.7	5.4	5.9	6.3	7.1	12.2	
Absolute sizing difference (ROI - Physician) (mm)									
Case	1	2	3	4	5	6	7	8	Average
DAS	0.2	-0.1	-0.1	0.1	0.8	0.0	-1.4	0.1	-0.1 ± 0.6
AD+MV	0.5	-0.2	0.0	1.5	2.2	2.4	0.8	0.6	1.0 ± 1.0
AD-Ex	0.8	0.8	1.0	-0.3	0.4	1.1	0.5	4.8	1.1 ± 1.5
AD-Ex+MV	0.7	0.8	0.1	0.2	0.9	1.9	0.4	4.6	1.2 ± 1.5
AD-Ex+GCF	0.7	0.8	0.9	2.1	0.4	1.5	0.3	4.7	1.4 ± 1.5

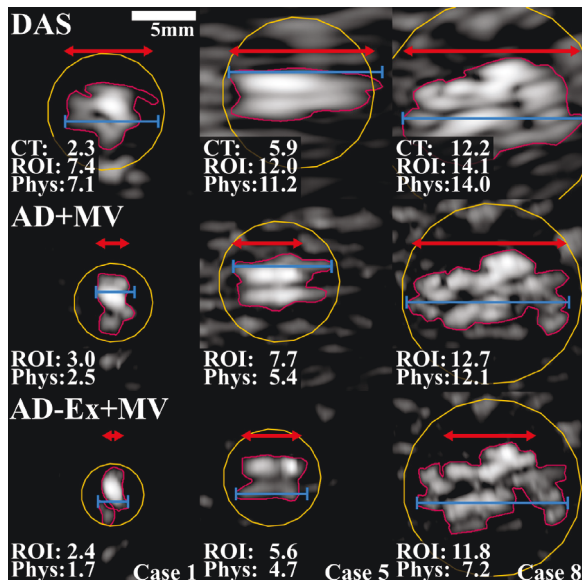


Fig. 8. DAS, AD+MV, and AD-Ex+MV 30 dB dynamic range b-mode images for Cases 1, 5, and 8. The measured CT size, size based on the ROI, and size measured by the physician are included for each case. The red dimension line with arrows shows the physician measured size, while the blue dimension line with tick marks shows the ROI measured size. In all cases, the physician measures a smaller dimension compared to the ROI method.

the base. It demonstrates the advantages of combining AD-Ex and MV, since they improve different aspects of the main lobe, and the combined method AD-Ex+MV carries over and improves both of these aspects. However, these simulations do not fully represent *in vivo* performance. For example, MV appears better than GCF in these simulations, while *in vivo* GCF demonstrates a significant improvement compared to MV. ADMIRE, and AD-Ex especially, generally reduce clutter to enhance the target while MV sharpens the target, which can result in different stone appearances based on content of the original image. This overall suggests that improving the clarity of the coherent target, such as by boosting coherence with GCF or decluttering with ADMIRE, is critical to improving sizing accuracy, rather than strictly improving resolution.

The cyst simulations show how ultrasound is overestimating hyper-echoic targets. In particular, overestimation occurs when we measure the targets at traditional b-mode dynamic ranges (30–50 dB or more). Since methods like MV narrow the main lobe mostly in the range of -20 to 0 dB, this results in marginal improvements measured at 30 dB. In comparison, measuring on an image with a dynamic range of 6 dB (i.e. at the FWHM) results in underestimation, and a much larger difference for MV compared to DAS. This implies that there may be a theoretical dynamic range where a method is more likely to measure a target correctly, though the *in vivo* examples indicate that this threshold would vary between methods, and potentially depend on the amplitude of the target versus the background. In the *in vivo* cases, DAS overestimates consistently at -30 dB, and mostly seems to be accurate between -6 and -12 dB. However, GCF and AD-Ex are substantially more variable for the same cases. Eventually this could potentially be a tool to improve ultrasound sizing accuracy, though it also follows that this threshold would have to be tuned for each method of interest.

For the sizing tasks *in vivo*, we continue to observe as in previous work that the ultrasound methods tend to overestimate size compared to CT. For these results, we provided both contour-based ROI selections to determine the size of the stone and the physician-based measurements. The ROI results showed that AD-Ex was able to attain comparable or slightly improved sizing performance compared to AD+MV, while also improving contrast ratio and gCNR. This is already valuable, as removing the need to run additional processing with MV saves computation time. However, the other advantage is that we can further process the results to attain even more improvements, and AD-Ex+MV was able to further reduce average error, managing an average of 10.8% compared to AD+MV with 27.8%. While perhaps seemingly modest compared to the already significant improvement compared to DAS, improving accuracy and consistency of those results will continue

to improve the chance of correctly sizing stones in borderline-surgery cases.

Despite the improvements to lateral performance that AD-Ex generates, there are several cases where after processing the axial dimension becomes the longest observed dimension, which may beg the question of why we continue to size with the lateral dimension. AD-Ex was designed to operate purely on those off-axis signals, so we do not expect improvements axially. Additionally, sizing targets along the axial dimension, especially hard coherent targets like kidney stones, is considered less reliable than along the lateral dimension. A primary factor for this is reverberation that can occur within the stone, which can produce an axial artifact underneath the stone often referred to as a comet trail. As a result, sonographers specifically aim to make sizing measurements along the lateral dimension when possible. It is possible that even if an acquisition was made in the perpendicular dimension, something similar would be observed, and that this effect is not due to the orientation of the stone. Though we cannot confirm with certainty why Cases 1, 2, and 6 take on the appearances that they do, there was an effort to use the CT acquisitions to align the ultrasound acquisitions such that the longest stone axis was along the lateral dimension.

There continues to be a large amount of variability between stones from different patients, and while AD-Ex+MV does reduce much of this variance, the root cause of this variability is not fully understood. We included both BMI and the depth of the stone for each case, however, there was no observable trend or correlation between those features and the observed error for any method. For example, Cases 1 and 2 are both small stones in low BMI patients at a relatively shallow depth, but the average error for DAS is vastly different (215.2% for Case 1 and 58.8% for Case 2). In comparison, Case 6 has by far the highest BMI and a moderately deep stone, but has relatively low error compared to the other cases. Especially for a preliminary study this small, there is simply not enough data to pick up trends since we are unable to differentiate which stones may be outliers.

However, there is one major conflict that remains to be resolved here, which is the differences between the ROI measurements and the physician measurements. In particular, the physician measurements were consistently smaller compared to the ROI-based measurements. This means that AD+MV generally measured more accurately compared to CT by the physician, but the AD-Ex methods tended to actually underestimate compared to CT in several cases. There are two reasons why this occurred. In some cases, in particular Case 8, the physician actually selected a small part of the overall stone with AD-Ex, resulting in a significant underestimation of the size. In other cases, as mentioned in the results section, the differences arise when peripheral content was or was not considered as part of the stone.

Using the contour map to choose the ROI resulted in fairly consistent choices, whereas there are several potential factors that could lead to increased variance in the physician measurements. For example, the sizing occurred on a 30 dB dynamic range image, which is much lower than would be used clinically. For DAS this does not make much of a difference, but for the AD-Ex images much of the background is suppressed. This can result in more disjointed-looking stone structures which may be interpreted by a physician differently. Additionally, the actual contours tend to pick up more peripheral content that may be low amplitude in the image and could be missed or ignored by the physician, while it would be included in the ROI workflow. Finally, there is always potential variability inherent to human nature. A physician is likely to be familiar with the behavior of stones in DAS images. However, since AD-Ex and other adaptive beamformers make significant qualitative changes to b-mode images, it is possible that the intuition and expertise of a physician built on reviewing DAS images may not be the best way to take advantage of these methods.

6. Conclusion

AD-Ex shows promise as a pre-processing step for improving sizing accuracy in ultrasound. On the ROI-based stone measurements, AD-Ex had an average error of 23.4%, slightly improving on the previous ADMIRE+MV which was 27.8%. However, the strength of AD-Ex is that we can perform additional post-processing, and in this work we have shown that post-processing with MV results in lowering the average error to 10.8%. Overall, AD-Ex+MV was able to demonstrate improved average error with lower variance between cases.

However, the included physician-based measurements show that there are still some issues that need to be resolved. For the physician, the higher contrast images of AD-Ex displayed on a lower dynamic range than is normal clinically made it difficult to choose the exact stone region, despite these cases being easier for the ROI-based method. This may imply that a workflow that leverages more information or even some amount of automatic assistance may be worth investigating in the future, which may allow us to increase physician measurement consistency.

Finally, while our data does suggest that there may be a theoretical dynamic range that would produce more accurate ultrasound sizing measurements, it will require a larger data set. If in the future such a data set was available, this would potentially be a great target for a deep learning approach as well, where a machine might be able to tease apart some interactions or parameters that a practitioner cannot, for example taking into account factors such as amount and type of surrounding tissue, transducer parameters, or even beamformer specific factors like general image quality.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Brett Byram has patent #US20190365345A1 pending to Vanderbilt University.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank the staff of the Vanderbilt University ACCRE computing resource and the VISE Physician-in-Residence Program. This work was supported by NIH, United States grants R01EB020040, R01HL156034, and S10OD023680-01.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.ultras.2023.106952>.

References

- [1] M.A. Lediju, G.E. Trahey, B.C. Byram, J.J. Dahl, Short-lag spatial coherence of backscattered echoes: Imaging characteristics, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 58 (7) (2011) 1377–1388, <http://dx.doi.org/10.1109/TUFFC.2011.1957>.
- [2] S. Schlunk, J. Tierney, M. George, P. Karve, R. Duddu, R.S. Hsi, B. Byram, Non-linear beamforming approaches for sizing and detecting large calcifications, in: 2017 IEEE International Ultrasonics Symposium, IUS, IEEE, 2017, pp. 1–4, <http://dx.doi.org/10.1109/ULTSYM.2017.8092597>, URL <http://ieeexplore.ieee.org/document/8092597/>.
- [3] R.S. Hsi, S.G. Schlunk, J.E. Tierney, K. Dei, R. Jones, M. George, P. Karve, R. Duddu, B.C. Byram, Feasibility of non-linear beamforming ultrasound methods to characterize and size kidney stones, *PLoS One* 13 (8) (2018) 1–14, <http://dx.doi.org/10.1371/journal.pone.0203138>.

- [4] J.E. Tierney, S.G. Schlunk, R. Jones, M. George, P. Karve, R. Duddu, B.C. Byram, R.S. Hsi, In vitro feasibility of next generation non-linear beamforming ultrasound methods to characterize and size kidney stones, *Urolithiasis* 47 (2) (2019) 181–188, <http://dx.doi.org/10.1007/s00240-018-1036-z>.
- [5] R. Jones, B.C. Byram, S. Schlunk, J.E. Tierney, R. Hsi, The impact of mid lag spatial coherence parameters on coherent target detection, in: N.V. Ruiter, B.C. Byram (Eds.), in: *Medical Imaging 2019: Ultrasonic Imaging and Tomography*, (no. March 2019) SPIE, 2019, p. 22, <http://dx.doi.org/10.1117/12.2514118>, URL <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10955/2514118/The-impact-of-mid-lag-spatial-coherence-parameters-on-coherent/10.1117/12.2514118.full>.
- [6] W. Shabana, R.O. Bude, J.M. Rubin, Comparison between color Doppler twinkling artifact and acoustic shadowing for renal calculus detection: An in vitro study, *Ultrasound Med. Biol.* 35 (2) (2009) 339–350, <http://dx.doi.org/10.1016/j.ultrasmedbio.2008.09.023>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0301562908004456>.
- [7] M. Abdel-Gawad, R.D. Kadasne, E. Elsobky, B. Ali-El-Dein, M. Monga, A prospective comparative study of color Doppler ultrasound with twinkling and noncontrast computerized tomography for the evaluation of acute renal colic, *J. Urol.* 196 (3) (2016) 757–762, <http://dx.doi.org/10.1016/j.juro.2016.03.175>, URL <http://www.jurology.com/doi/10.1016/j.juro.2016.03.175>.
- [8] F.L. Coe, The natural history of calcium urolithiasis, *JAMA: J. Am. Med. Assoc.* 238 (14) (1977) 1519, <http://dx.doi.org/10.1001/jama.1977.03280150089036>, URL <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.1977.03280150089036>.
- [9] A. Ueno, T. Kawamura, A. Ogawa, H. Takayasu, Relation of spontaneous passage of ureteral calculi to size, *Urology* 10 (6) (1977) 544–546, [http://dx.doi.org/10.1016/0090-4295\(77\)90097-8](http://dx.doi.org/10.1016/0090-4295(77)90097-8).
- [10] S. Ulasun, Z. Koc, N. Tokmak, Accuracy of sonography for detecting renal stone: Comparison with CT, *J. Clin. Ultrasound* 35 (5) (2007) 256–261, <http://dx.doi.org/10.1002/jcu.20347>, URL <http://doi.wiley.com/10.1002/jcu.20347>.
- [11] D. Unal, E. Yeni, M. Karaoglanoglu, A. Verit, O.F. Karatas, Can conventional examinations contribute to the diagnostic power of unenhanced helical computed tomography in urolithiasis? *Urol. Int.* 70 (1) (2003) 31–35, <http://dx.doi.org/10.1159/000067702>, URL <https://www.karger.com/Article/FullText/67702>.
- [12] A.A. Ray, D. Ghiculete, K.T. Pace, R.J.D. Honey, Limitations to ultrasound in the detection and measurement of urinary tract calculi, *Urology* 76 (2) (2010) 295–300, <http://dx.doi.org/10.1016/j.urol.2009.12.015>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0090429509030799>.
- [13] V. Ganesan, S. De, D. Greene, F.C.M. Torricelli, M. Monga, Accuracy of ultrasonography for renal stone detection and size determination: Is it good enough for management decisions? *BJU Int.* 119 (3) (2017) 464–469, <http://dx.doi.org/10.1111/bju.13605>, URL <http://doi.wiley.com/10.1111/bju.13605>.
- [14] K.A.B. Fowler, J.A. Locken, J.H. Duchesne, M.R. Williamson, US for detecting renal calculi with nonenhanced CT as a reference standard, *Radiology* 222 (1) (2002) 109–113, <http://dx.doi.org/10.1148/radiol.2221010453>, URL <http://pubs.rsna.org/doi/10.1148/radiol.2221010453>.
- [15] B. Dunmire, F.C. Lee, R.S. Hsi, B.W. Cunitz, M. Paun, M.R. Bailey, M.D. Sorensen, J.D. Harper, Tools to improve the accuracy of kidney stone sizing with ultrasound, *J. Endourol.* 29 (2) (2015) 147–152, <http://dx.doi.org/10.1089/end.2014.0332>, URL <http://www.liebertpub.com/doi/10.1089/end.2014.0332>.
- [16] K.M. Sternberg, B. Eisner, T. Larson, N. Hernandez, J. Han, V.M. Pais, Ultrasonography significantly overestimates stone size when compared to low-dose, noncontrast computed tomography, *Urology* 95 (2016) 67–71, <http://dx.doi.org/10.1016/j.urol.2016.06.002>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0090429516302710>.
- [17] B. Dunmire, J.D. Harper, B.W. Cunitz, F.C. Lee, R. Hsi, Z. Liu, M.R. Bailey, M.D. Sorensen, Use of the acoustic shadow width to determine kidney stone size with ultrasound, *J. Urol.* 195 (1) (2016) 171–177, <http://dx.doi.org/10.1016/j.juro.2015.05.111>.
- [18] W. King, C. Kimme-Smith, J. Winter, Renal stone shadowing: An investigation of contributing factors, *Radiology* 154 (1) (1985) 191–196, <http://dx.doi.org/10.1148/radiology.154.1.3880605>, URL <http://pubs.rsna.org/doi/10.1148/radiology.154.1.3880605>.
- [19] C.L. Moore, L. Scoult, Sonography first for acute flank pain? *J. Ultrasound Med.* 31 (11) (2012) 1703–1711, <http://dx.doi.org/10.7863/jum.2012.31.11.1703>.
- [20] P.C. May, Y. Haider, B. Dunmire, B.W. Cunitz, J. Thiel, Z. Liu, M. Bruce, M.R. Bailey, M.D. Sorensen, J.D. Harper, Stone-mode ultrasound for determining renal stone size, *J. Endourol.* 30 (9) (2016) 958–962, <http://dx.doi.org/10.1089/end.2016.0341>, URL <http://www.liebertpub.com/doi/10.1089/end.2016.0341>.
- [21] R.C. Smith, M. Varanelli, Diagnosis and management of acute ureterolithiasis, *Am. J. Roentgenol.* 175 (1) (2000) 3–6, <http://dx.doi.org/10.2214/ajr.175.1.1750003>, URL <http://www.ajronline.org/doi/10.2214/ajr.175.1.1750003>.
- [22] R.C. Smith, M. Verga, S. McCarthy, A.T. Rosenfield, Diagnosis of acute flank pain: Value of unenhanced helical CT, *Am. J. Roentgenol.* 166 (1) (1996) 97–101, <http://dx.doi.org/10.2214/ajr.166.1.8571915>, URL <http://www.ajronline.org/doi/10.2214/ajr.166.1.8571915>.
- [23] S. Schlunk, B. Byram, Expanded beamforming models for high dynamic range scenarios, in: 2020 IEEE International Ultrasonics Symposium, Vol. 2020-Septe, no. Mv, IUS, IEEE, 2020, pp. 1–4, <http://dx.doi.org/10.1109/IUS46767.2020.9251520>, URL <https://ieeexplore.ieee.org/document/9251520/>.
- [24] S. Schlunk, B.C. Byram, Using ADMIRE to improve minimum variance performance in the presence of reverberation clutter, in: N.V. Ruiter, B.C. Byram (Eds.), in: *Medical Imaging 2021: Ultrasonic Imaging and Tomography*, (no. 1) SPIE, 2021, p. 1, <http://dx.doi.org/10.1117/12.2583314>, URL <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11602/2583314/Using-ADMIRE-to-improve-minimum-variance-performance-in-the-presence/10.1117/12.2583314.full>.
- [25] S. Schlunk, B. Byram, Combining ADMIRE and MV to improve image quality, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 69 (9) (2022) 2651–2662, <http://dx.doi.org/10.1109/TUFFC.2022.3194548>, URL <https://ieeexplore.ieee.org/document/9844004/>.
- [26] I.K. Holford, F. Gran, J.A. Jensen, Broadband minimum variance beamforming for ultrasound imaging, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 56 (2) (2009) 314–325, <http://dx.doi.org/10.1109/TUFFC.2009.1040>.
- [27] J.-F. Synnevåg, A. Austeng, S. Holm, Adaptive beamforming applied to medical ultrasound imaging, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 54 (8) (2007) 1606–1613.
- [28] P.-c. Li, M.-L. Li, Adaptive imaging using the generalized coherence factor, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 50 (2) (2003) 128–141, <http://dx.doi.org/10.1109/TUFFC.2003.1182117>.
- [29] O.M.H. Rindal, A. Rodriguez-Molares, A. Austeng, The Dark Region artifact in adaptive ultrasound beamforming, in: *IEEE International Ultrasonics Symposium, IUS*, 2017, pp. 1–4.
- [30] S. Schlunk, K. Dei, B. Byram, Iterative model-based beamforming for high dynamic range applications, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 68 (3) (2021) 482–493, <http://dx.doi.org/10.1109/TUFFC.2020.3012165>, URL <https://ieeexplore.ieee.org/document/9149929/>.
- [31] B. Byram, K. Dei, J. Tierney, D. Dumont, A model and regularization scheme for ultrasonic beamforming clutter reduction, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 62 (11) (2015) 1913–1927, <http://dx.doi.org/10.1109/TUFFC.2015.007004>.
- [32] K. Dei, B. Byram, The impact of model-based clutter suppression on cluttered, aberrated wavefronts, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 64 (10) (2017) 1450–1464.
- [33] B. Byram, M. Jakovljevic, Ultrasonic multipath and beamforming clutter reduction: A chirp model approach, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 61 (3) (2014) 428–440, <http://dx.doi.org/10.1109/TUFFC.2014.2928>.
- [34] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 67 (2) (2005) 301–320.
- [35] R.J. Tibshirani, J. Taylor, Degrees of freedom in Lasso problems, *Ann. Statist.* 40 (2) (2012) 1198–1232, <http://dx.doi.org/10.1214/12-AOS1003>.
- [36] B. Yang, A study of inverse short-time Fourier transform, in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 3541–3544.
- [37] C. Khan, K. Dei, B. Byram, A GPU-based implementation of ADMIRE, in: 2019 IEEE International Ultrasonics Symposium, Vol. 2019-October, IUS, IEEE, 2019, pp. 1501–1504, <http://dx.doi.org/10.1109/ULTSYM.2019.8925842>, URL <https://ieeexplore.ieee.org/document/8925842/>.
- [38] C. Khan, K. Dei, S. Schlunk, K. Ozgun, B. Byram, Real-time, simultaneous DAS, ADMIRE, and SLSC imaging using GPU-based processing, in: 2020 IEEE International Ultrasonics Symposium, IUS, IEEE, 2020, pp. 1–4, <http://dx.doi.org/10.1109/IUS46767.2020.9251587>, URL <https://ieeexplore.ieee.org/document/9251587/>.
- [39] S. Schlunk, K. Dei, B. Byram, Iterative ADMIRE for high dynamic range B-mode, in: *IEEE International Ultrasonics Symposium*, Vol. 2018-October, IUS, IEEE, 2018, pp. 1–4, <http://dx.doi.org/10.1109/ULTSYM.2018.8580209>.
- [40] K. Dei, J. Tierney, B. Byram, Plane wave image quality improvement using ADMIRE algorithm, in: *IEEE International Ultrasonics Symposium*, Vol. 2016-Novem, no. Mv, IUS, IEEE, 2016, pp. 0–3, <http://dx.doi.org/10.1109/ULTSYM.2016.7728891>.
- [41] J.A. Jensen, Field: A Program for Simulating Ultrasound Systems, in: *Paper Presented At the 10th Nordic-Baltic Conference on Biomedical Imaging Published in Medical & Biological Engineering and Computing*, Vol. 34, 1996, pp. 351–353.
- [42] J.A. Jensen, N.B. Svendsen, Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 39 (1992) 262–267.
- [43] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, M. Fink, Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 56 (3) (2009) 489–506, <http://dx.doi.org/10.1109/TUFFC.2009.1067>.
- [44] A. Rodriguez-Molares, O.M.H. Rindal, J. D'Hooge, S.-E. Måsøy, A. Austeng, H. Torp, The Generalized Contrast-to-Noise Ratio, in: *IEEE International Ultrasonics Symposium, IUS*, 2018, pp. 1–4.
- [45] A. Rodriguez-Molares, O.M.H. Rindal, J. D'Hooge, S.-E. Masoy, A. Austeng, M.A. Lediju Bell, H. Torp, The generalized contrast-to-noise ratio: A formal definition for lesion detectability, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 67 (4) (2020) 745–759, <http://dx.doi.org/10.1109/TUFFC.2019.2956855>, URL <https://ieeexplore.ieee.org/document/8918059/>.