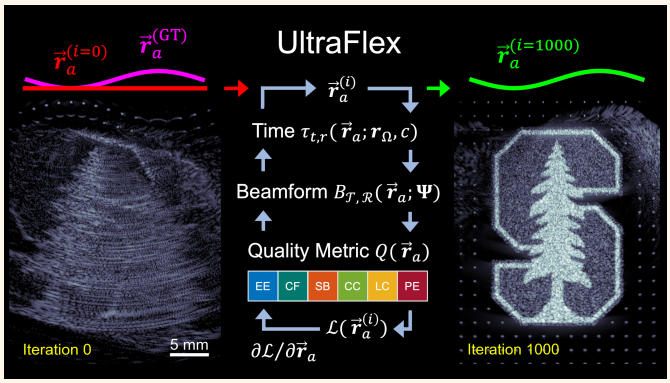
UltraFlex: Iterative Model-Based Ultrasonic Flexible-Array Shape Calibration

Benjamin N. Frey[1], Dongwoon Hyun[2,3], Walter Simson[2], Louise Zhuang[4], Hoda S. Hashemi[2], Martin Schneider[2], and Jeremy J. Dahl[2]

***Abstract*— UltraFlex is an iterative model-based ultrasonic flexible-array shape calibration framework that uses automatic differentiation. This work evaluates array-shape-calibration model performance while examining multiple image quality metrics: speckle brightness, envelope entropy, coherence factor, lag-one coherence, common-midpoint correlation coefficient, and common-midpoint phase error. The accuracy of these image quality metrics was evaluated on simulated phantoms using a variety of array shapes. Experimental phantom and in vivo liver datasets were also investigated using transducers with known geometries. While speckle brightness, envelope entropy, and coherence factor enable model convergence under many conditions, lag-one coherence, common-midpoint correlation coefficient, and common-midpoint phase error enable more accurate element position estimations and improved visual ultrasound image focusing quality. Furthermore, the models based on the common-midpoint correlation coefficient and phase-error quality metrics are the most robust against additive white noise while achieving median mean Euclidean errors (MEEs) of 3.7 µm for simulation, 29.7 µm for phantom, and 69.0 µm for in vivo liver data. These array shape calibration results show promise for the current and future development of experimental flexible- and wearable-ultrasonic arrays.**



***Index Terms*— flexible arrays, wearable ultrasound, acoustic imaging, ultrasound autofocusing**

I. Introduction

FOR decades, medical ultrasound imaging has been confined to conventional rigid-array transducers. While these conventional transducers have provided immeasurable benefit to ultrasound imaging, the recent development of flexible-array ultrasound transducers opens the door to new applications for ultrasound, including continuous monitoring and wearable ultrasound technologies [1]–[5]. Flexible-array ultrasound systems provide a conformable aperture that can accommodate the contoured surface of an imaging subject. Furthermore, some flexible-array ultrasound systems can be used in a wearable configuration to alleviate the need for continuous ultrasound navigation, reducing operator dependency [1].

For example, rodent-wearable flexible-array transducers have previously been investigated with the goal of neural activity monitoring [2]. Human-wearable flexible-array device studies have been designed for assessing cardiac dysfunction via cardiac output, ejection fraction, and stroke volume [3]. Another application area of wearable flexible-array ultrasound systems can be found in therapeutics. Wound healing with focused ultrasound has demonstrated protein stimulation in dermal and epidermal layers of diabetic rats [4]. Neuromodulation using a prototype wearable flexible-array has also been proposed [5]. Despite the benefits that flexible-array ultrasound systems provide, image reconstruction and therapeutic focusing are compromised without knowing the pose of each transducer element (that is, orientation and position in $\mathbb{R}^3$).

Two primary approaches have been used to estimate unknown element poses. The first class involves external mechanical (e.g., fiber-optic strain [6], resistive-strain, etc.) or optical [7], [8] tracking-based analysis of array deformation. Chen et al. [6] used a near-infrared fiber-optic reflectometer device embedded within a custom flexible-array transducer to achieve position errors of 400 µm and 421 µm in the y–z (lateral) plane for a convex and sinusoid array shape, respectively. China et al. [8] passively tracked infrared-reflecting spheres affixed to the flexible array to achieve position errors of 500 ± 290 µm (CIRS phantom), 540 ± 350 µm (deformable phantom), and 360 ± 240 µm (cadaveric specimen). These approaches require external equipment that adds bulk to the transducer, additional

Manuscript received May 18, 2025, accepted [Date], and published [Date]. This research was supported in part by the National Institute of Biomedical Imaging and Bioengineering through Grants R01-EB027100 and K99-EB032230.

Departments of Applied Physics[1], Radiology[2], and Electrical Engineering[4], Stanford University, Stanford, CA 94305 USA.

[3]Siemens Healthineers, Palo Alto, CA 94304 USA.

B. N. Frey (e-mail: benfrey@stanford.edu) and J. J. Dahl (e-mail: jjdahl@stanford.edu) are the corresponding authors.

Highlights

- **UltraFlex is a model-based flexible-array shape calibration framework that uses iterative optimization updates through automatic differentiation to enable lower position estimation error compared to previous models.**
- **Model performance is evaluated using envelope entropy, speckle brightness, lag-one coherence, coherence factor, correlation coefficient, and phase-error focusing quality metrics.**
- **These iterative model results demonstrate promise for current and future hardware development of ultrasonic wearable and flexible arrays that utilize ultrasound-autofocusing software.**

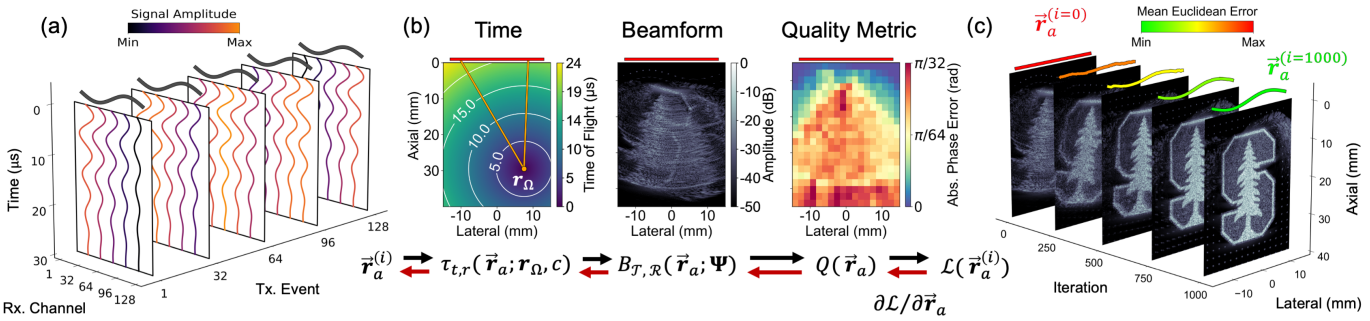


Fig. 1. (a) Acquisition of the multistatic dataset $\boldsymbol{\Psi}$. (b) The iterative aperture shape estimation process begins by computing the bi-directional time-of-flight $\boldsymbol{\tau_{t,r}}$. From these delays, the IQ data is beamformed $\boldsymbol{B}$ using a synthetic aperture approach (the image above has been envelope-detected and log-compressed). An image quality metric $\boldsymbol{Q}$ is evaluated, which is then used as part of an objective function $\boldsymbol{\mathcal{L}}$. The objective function is differentiated with respect to the element positions $\vec{\boldsymbol{r}}_{\boldsymbol{a}}$, and the red arrows represent the error that is backpropagated through the model to update the array element positions at iteration $\boldsymbol{i}$. (c) Visualization of B-mode reconstruction over iterations, starting with the aperture shape before the first iteration and after one thousand iterations. The color of the aperture loosely represents the deviation of the aperture shape from the ground truth, where red and green represent more and less deviation, respectively.

complexity to clinical translation, or both.

The second class of approaches involves intrinsic determination of element pose via algorithms that rely only on the channel data collected from the imaging system. These approaches include methods based on assumed array geometries [9]–[11], derivative-free black-box optimization (e.g. simulated annealing [12], [13] or exhaustive sampling [14], [15]), deep learning methods [16], [17], and direct gradient-based optimization [7], [18]. Some approaches within this class assess an image quality metric such as the maximum lag-one spatial coherence [11], image contrast [12], envelope amplitude-variance region-of-interest sharpness [13], phase variance [14], short-lag spatial coherence [15], and image entropy [7], [18]. Ingram et al. [14] achieved root mean square errors in element positions of 0.18 λ (110.9 μm) and 0.4 λ (246.4 μm) for simulated and experimental data, respectively. Omidvar et al. [15] achieved mean Euclidean errors (MEEs) lower than 0.1 λ (43.1 μm) and 1.4 λ (603.9 μm) for simulated and experimental data, respectively. Noda et al. [16] achieved an average MEE of 860 μm and 1,110 μm for simulation and in vivo test data, respectively, using a deep learning approach. Noda et al. [18] reported average MEEs between 40 and 54 μm across three array geometries for simulation data, and between 8.6 and 37 μm across three array geometries for phantom data while using a direct gradient-based approach with an aperture size of only 20 elements in each case. Huang et al. [7] built upon [18] with coarse initialization from an optical tracker while achieving a post-iterative-model MEE of 148.8 μm.

As mentioned previously, methods within the second class of approaches may be derivative-free or involve gradient backpropagation via automatic differentiation frameworks (e.g., JAX [19], PyTorch, and TensorFlow). Conventional deep learning-based methods utilize gradient backpropagation to update the weights and biases of a multilayer neural network. A direct gradient-based optimization approach may similarly utilize an automatic differentiation framework with a differentiable objective function to iteratively update model assumptions (e.g., array element poses) [18].

In Noda et al. [18], an array shape estimation is made using a linear combination of orthogonal functions. Based on this array shape, timing delays are calculated, and image focus is evaluated by calculating the beamformed envelope entropy. The errors are backpropagated through the beamformer to update the combination of orthogonal functions. Our method differs from Noda et al. [18] in that (a) we do not constrain our estimated shape to a combination of orthogonal functions (our only assumption is that the number of elements and the pitch of the array are known), (b) we preferentially use a differentiable objective function that is based on common-midpoint signals and is not subject to errors resulting from natural signal decorrelation or variations in signal amplitude, (c) we account for the orientation of individual elements, and (d) we validate our model on in vivo liver data. Here, we extend our previous work in Hyun et al. [20] to rigorously evaluate our UltraFlex framework across various quality metric-based objective functions, update the reconstruction model to account for curved surfaces, and examine the performance on phantoms and in vivo data.

II. UltraFlex Framework

A. Differentiable Image Reconstruction

Pulse-echo ultrasound imaging utilizes transmit and receive beamforming to focus pressure signals (i.e., to spatially localize them). Let $\boldsymbol{r}_a$ denote the spatial position of the a-th array element in the x–z (elevational) plane. Assuming a constant speed of sound c, the time-of-flight from the element to the coordinate $\boldsymbol{r}_\Omega$ in the imaging field-of-view Ω is

$$t_d(\boldsymbol{r}_a; \boldsymbol{r}_\Omega, c) = \frac{1}{c} \left\| \boldsymbol{r}_\Omega - \boldsymbol{r}_a \right\|_2 . \tag{1}$$

Elements with spatial extent exhibit directivity, which can be modeled using an acceptance cone mask based on f-number. First, we transform $\boldsymbol{r}_\Omega$ into the element's frame of reference using a translation and rotation:

$$\boldsymbol{r}'_\Omega = \begin{bmatrix} x'_\Omega \\ z'_\Omega \end{bmatrix} = \begin{bmatrix} \cos\theta(\boldsymbol{r}_a) & -\sin\theta(\boldsymbol{r}_a) \\ \sin\theta(\boldsymbol{r}_a) & \cos\theta(\boldsymbol{r}_a) \end{bmatrix} \begin{bmatrix} x_\Omega - x_a \\ z_\Omega - z_a \end{bmatrix}, \tag{2}$$

where the element orientation θ_a is inferred from neighboring elements as

$$\theta(\boldsymbol{r}_a) = -\tan^{-1}\left((z_{a+1} - z_{a-1}) / (x_{a+1} - x_{a-1}) \right). \tag{3}$$

Then, the acceptance cone mask of the a-th element is

$$h_a(\boldsymbol{r}_\Omega) = |x'_\Omega| \leq \frac{F}{2} z'_\Omega. \tag{4}$$

Given the set of all element positions $\vec{\boldsymbol{r}}_a$, the round-trip time-of-flight from the t-th transmit element to $\boldsymbol{r}_\Omega$ and back to the r-th receive element is simply $\tau_{t,r}(\vec{\boldsymbol{r}}_a) = t_d(\boldsymbol{r}_t) + t_d(\boldsymbol{r}_r)$.

Transmit and receive beamforming are applied as follows. Let $\mathcal{A}$ be the aperture element indices $\{1, 2, \ldots, N\}$. For a transmit subaperture $\mathcal{T} \subseteq \mathcal{A}$ and receive subaperture $\mathcal{R} \subseteq \mathcal{A}$, the beamformed signal is

$$B_{\mathcal{T},\mathcal{R}}(\vec{\boldsymbol{r}}_a; \boldsymbol{r}_\Omega, \boldsymbol{\Psi}) = \sum_{t \in \mathcal{T}} \sum_{r \in \mathcal{R}} \boldsymbol{\Psi}_{t,r}(\tau_{t,r}) \exp(j2\pi f_d \tau_{t,r}) h_t h_r. \tag{5}$$

where $\boldsymbol{\Psi}$ is the pulse-echo response matrix, the abbreviation $\tau_{t,r} = \tau_{t,r}(\vec{\boldsymbol{r}}_a)$ has been made, and f_d is the demodulation frequency. For notational simplification, B will often be expressed without a position $\boldsymbol{r}_\Omega$ in the image domain or the pulse-echo matrix $\boldsymbol{\Psi}$ because these arguments are constant across all model iterations. Because each step is differentiable, the entire image reconstruction process is differentiable.

B. Array Shape Calibration

Automatic differentiation frameworks such as JAX [19], PyTorch, and Tensorflow can offer model representation and optimization by storing differentiable operations in a computational graph. Similarly to training neural networks, we use automatic differentiation to minimize an objective function, where gradient backpropagation is used to update unknown model parameters (in this case, the element positions of the array). The objective function errors are backpropagated as gradients through a physical model, in this case, the differentiable image reconstruction process, rather than a set of neural network layers, to directly update the estimated element positions.

The entire iterative flexible array shape estimation process is shown in Fig. 1. The flexible array will be referred to as an imaging aperture. First (Fig. 1a), a full synthetic aperture (FSA) transmission sequence is utilized to collect a pulse-echo dataset $\boldsymbol{\Psi}$ from an unknown aperture shape. During an FSA transmission sequence, only one element is activated during a transmission event while all the other elements receive the back-scattered echoes. This transmission event is repeated for each element of the aperture, forming a multistatic dataset (i.e., a tensor with dimensions of time by number of transmit channels by number of receive channels). The acquired RF data is demodulated to extract the in-phase and quadrature (IQ) baseband components. Next (Fig. 1b), an initial assumption is made about the aperture geometry (e.g., the aperture shape is a linear array) given a known pitch. From this aperture shape, element timing delays $\tau_{t,r}(\vec{\boldsymbol{r}}_a)$ are geometrically calculated. Finally, the IQ data is beamformed over a set of image domain positions $\vec{\boldsymbol{r}}_\Omega$, and a quality metric Q is calculated. The metric is used as a term in an optimizable objective function, and the objective error is backpropagated through the quality calculation, the beamforming process, and the timing-delay calculation to update the estimated element positions of the aperture. This process can be repeated iteratively, where *all* estimated element positions $\vec{\boldsymbol{r}}_a$ from the current iteration i are used to initialize the next iteration $i+1$. The update rule is

$$\vec{\boldsymbol{r}}_a^{(i+1)} = \vec{\boldsymbol{r}}_a^{(i)} - \gamma \frac{\partial \mathcal{L}}{\partial \vec{\boldsymbol{r}}_a} \left(\vec{\boldsymbol{r}}_a^{(i)} \right), \tag{6}$$

where γ is a step size parameter, and $\partial \mathcal{L}(\vec{\boldsymbol{r}}_a^{(i)}) / \partial \vec{\boldsymbol{r}}_a$ denotes the gradient of the objective function $\mathcal{L}$ with respect to the element positions $\vec{\boldsymbol{r}}_a$, evaluated at iteration i. In other words, element positions are updated in the direction of the negative gradient of the objective function.

C. Quality Metrics

Seven image quality metrics from the literature are adapted to the UltraFlex framework for evaluation. For the speckle brightness, envelope entropy, coherence factor, and lag-one coherence metrics below, each metric expression is evaluated using the full transmit $\mathcal{T} = \mathcal{A}$ and receive $\mathcal{R} = \mathcal{A}$ apertures (i.e., $B_{\mathcal{T},\mathcal{R}} = B_{\mathcal{A},\mathcal{A}}$). For the common-midpoint correlation coefficient and common-midpoint phase-error metrics, each metric expression is evaluated using a pair of transmit ($\mathcal{T}_\alpha, \mathcal{T}_\beta \subset \mathcal{A}$) and receiver ($\mathcal{R}_\alpha, \mathcal{R}_\beta \subset \mathcal{A}$) subapertures.

1) Speckle Brightness: Image speckle brightness is defined as the magnitude of the coherent sum of images within a region of interest. Nock et al. [21] introduced the speckle brightness quality metric as part of a method for correcting unknown phase aberrations. The criterion can be expressed as

$$Q_{\text{SB}}(\vec{\boldsymbol{r}}_a) = |B_{\mathcal{A},\mathcal{A}}(\vec{\boldsymbol{r}}_a)|. \tag{7}$$

2) Envelope Entropy: Envelope entropy is a measure of image disorder or defocusing. When the envelope of a beamformed signal has high entropy, the image has more distortion due to malformed PSF functions. Noda et al. [18] adapted this focusing criterion for ultrasonic flexible-array shape estimation from applications in inverse synthetic-aperture radar [22]. The metric can be expressed as

$$Q_{\text{EE}}(\vec{\boldsymbol{r}}_a) = -B_{\text{env}}(\vec{\boldsymbol{r}}_a) \log_2 B_{\text{env}}(\vec{\boldsymbol{r}}_a), \tag{8}$$

where B_{env} represents the normalized envelope-detected signal over the sum of all domain pixels $\vec{\boldsymbol{r}}_\Omega$:

$$B_{\text{env}}(\vec{\boldsymbol{r}}_a) = \frac{|B_{\mathcal{A},\mathcal{A}}(\vec{\boldsymbol{r}}_a)|}{\sum_{v=1}^{N_\Omega} |B_{\mathcal{A},\mathcal{A}}(\vec{\boldsymbol{r}}_a)|}.$$

3) Coherence Factor: Mallart and Fink [23] introduced a focusing criterion, also known as the coherence factor, as a measure of focusing quality that is optimized with perfect focusing and decreases under increasing aberration. The coherence factor is a ratio between the coherent and incoherent beamformed signal sums across the receive aperture. The metric can be expressed as

$$Q_{\text{CF}}(\vec{\boldsymbol{r}}_a) = \frac{\left|\sum_{r=1}^{N_r} B_{\mathcal{A},r}\right|}{\sum_{r=1}^{N_r} |B_{\mathcal{A},r}|}, \tag{9}$$

where the abbreviation $B_{\mathcal{A},r} = B_{\mathcal{A},r}(\vec{\boldsymbol{r}}_a)$ refers to the transmit-beamformed receive-channel signals that have been time delayed, and N_r is the number of receive channels.

4) Lag-One Coherence: Lag-based spatial coherence describes the covariance between receiver-domain signals received at an element and its neighbor m elements (lags) away [23]. Lag-one coherence is a specific case of lag-based spatial coherence that only includes the first lag ($m = 1$) and was introduced as a quality metric by Long et. al [24]. The metric can be expressed as

$$Q_{\text{LC}}(\vec{\boldsymbol{r}}_a) = \frac{\Re\left\{\sum_{r=1}^{N_r-1} B_{\mathcal{A},r} B^*_{\mathcal{A},r+1}\right\}}{\sqrt{\sum_{r=1}^{N_r-1} |B_{\mathcal{A},r}|^2 |B_{\mathcal{A},r+1}|^2}}, \tag{10}$$

5) Common-Midpoint Correlation Coefficient (CMCC): The common-midpoint correlation coefficient (CMCC) is similar to the lag-one coherence quality metric but is applied in the common-midpoint domain. Based on Rachlin [25] and Ng et al. [26], signals correctly beamformed from common midpoint apertures theoretically should be the same (or otherwise highly correlated). The CMCC metric describes the correlation between two pairs of transmit and receive subapertures that share a common midpoint in the aperture domain. For example, apertures $\mathcal{T}_\alpha \subset \mathcal{T}$ and $\mathcal{R}_\alpha \subset \mathcal{R}$ may represent transmit and receive subapertures, respectively, each composed of N elements. Suppose that subapertures $\mathcal{T}_\beta \subset \mathcal{T}$ and $\mathcal{R}_\beta \subset \mathcal{R}$ are another pair of transmit and receive apertures, respectively, that are lag-one neighbors (one element separation) with $\mathcal{T}_\alpha$ and $\mathcal{R}_\alpha$ and share the same physical midpoint in the aperture domain [25], [26]. The correlation between $B_\alpha = B_{\mathcal{T}_\alpha,\mathcal{R}_\alpha}(\vec{\boldsymbol{r}}_a)$ and $B_\beta = B_{\mathcal{T}_\beta,\mathcal{R}_\beta}(\vec{\boldsymbol{r}}_a)$ is an example of the CMCC focusing criterion and is expressed as

$$Q_{\text{CC}}(\vec{\boldsymbol{r}}_a) = \frac{\Re\left\{\left\langle B_\alpha B^*_\beta \right\rangle\right\}}{\sqrt{\left\langle |B_\alpha|^2 \right\rangle \left\langle |B_\beta|^2 \right\rangle}}. \tag{11}$$

Here, the numerator represents the squared real part of the cross-correlation, while the denominator provides normalization by the product of the individual auto-correlations. Angle brackets represent the correlation over a spatial kernel.

6) Common-Midpoint Phase Error (CMPE): Common-midpoint phase error was introduced as an objective function by Simson et al. [27] and describes the phase error between two lag-one pairs of transmit and receive subapertures that share a common midpoint in the aperture domain. Similar to the CMCC metric, the phase error between $B_\alpha = B_{\mathcal{T}_\alpha,\mathcal{R}_\alpha}(\vec{\boldsymbol{r}}_a)$ and $B_\beta = B_{\mathcal{T}_\beta,\mathcal{R}_\beta}(\vec{\boldsymbol{r}}_a)$ can be expressed as

$$\Delta\phi_{\alpha,\beta}(\vec{\boldsymbol{r}}_a) = \angle\left[B_\alpha B^*_\beta\right],$$

and the quality metric for phase error is

$$Q_{\text{PE}}(\vec{\boldsymbol{r}}_a) = \left|\Delta\phi_{\alpha,\beta}(\vec{\boldsymbol{r}}_a)\right|. \tag{12}$$

The CMPE metric is the complex angle of the CMCC. The CMPE is computed from all lag-one pairs of transmit and receive subapertures, where $\alpha \subset A$ and $\beta \subset B$. Furthermore, a filtered version of the CMPE metric can be introduced to improve robustness against noise by selecting CMPE values at pixels with CMCC values $\geq \rho_{\text{thresh}}$. This ensures that only regions with sufficiently high cross-correlation contribute to the final metric. The filtered CMPE is defined as

$$Q_{\text{PEF}}(\vec{\boldsymbol{r}}_a) = \left|\Delta\phi_{\alpha,\beta}(\vec{\boldsymbol{r}}_a) \cdot \mathbb{1}(Q_{\text{CC}}(\vec{\boldsymbol{r}}_a) \geq \rho_{\text{thresh}})\right|, \tag{13}$$

where $\mathbb{1}(\cdot)$ is the indicator function and selects a subset of values returned by $\Delta\phi_{\alpha,\beta}(\vec{\boldsymbol{r}}_a)$. For the remainder of the paper, the CMPE and CMCC quality metrics will be referred to as phase error and correlation coefficient, respectively.

D. Objective Function

TABLE I
LEARNING RATES AND REDUCTIONS

Quality Metric	LR	Reduction Expression
Envelope Entropy	$1 \cdot 10^{-7}$	$\mathcal{L}_1 = \mathbb{E}\{5 \cdot 10^{-2} \cdot Q_{\text{EE}}\}$
Coherence Factor	$1 \cdot 10^{-7}$	$\mathcal{L}_1 = -\mathbb{E}\{Q_{\text{CF}}\}$
Speckle Brightness	$1 \cdot 10^{-7}$	$\mathcal{L}_1 = -\mathbb{E}\{2 \cdot 10^{-5} \cdot Q_{\text{SB}}\}$
Correlation Coefficient	$1 \cdot 10^{-7}$	$\mathcal{L}_1 = -\mathbb{E}\{10^{-2} \cdot \tanh^{-1}(Q_{\text{CC}})\}$
Lag-One Coherence	$7 \cdot 10^{-7}$	$\mathcal{L}_1 = -\mathbb{E}\{\tanh^{-1}(Q_{\text{LC}})\}$
Phase Error (Filtered)	$1 \cdot 10^{-7}$	$\mathcal{L}_1 = \mathbb{E}\{\ln(1 + (10^2 \cdot Q_{\text{PEF}})^2)\}$
Phase Error	$1 \cdot 10^{-7}$	$\mathcal{L}_1 = \mathbb{E}\{\ln(1 + (10^2 \cdot Q_{\text{PE}})^2)\}$

The objective function used for optimization incorporates one of the previously defined image quality metrics with regularization. The general expression for the objective function is given as

$$\begin{aligned}\mathcal{L}(\vec{\boldsymbol{r}}_a) = {} & w_1 \cdot \tanh\left[\mathcal{L}_1(Q(\vec{\boldsymbol{r}}_a))\right] \\ & + w_2 \cdot R_{\text{TV}}(\vec{\boldsymbol{r}}_a) \\ & + w_3 \cdot R_{\text{dz/dx}}(\vec{\boldsymbol{r}}_a),\end{aligned} \tag{14}$$

where w_j are weights applied to each term, and $\mathcal{L}_1$ is the reduction (transformation) of a quality metric. Learning rates and reduction expressions for each quality metric were empirically tuned for optimal performance within 1,000 iterations (as detailed in Table I). The expectation operator denotes the mean across all pixels and, in the case of the correlation coefficient and phase-error metrics, across all subaperture pairs. Speckle

brightness and envelope entropy were linearly scaled before averaging. Lag-one coherence and common-midpoint correlation coefficient metrics used a hyperbolic inverse tangent transform to emphasize values near the boundaries of (-1,1). Phase-error metrics were squared to penalize large errors and then passed through the natural logarithm of one plus the input for numerical stability. Negative signs in reduction expressions indicate metrics maximized by correct element positioning, unsigned expressions indicate metrics minimized by correct element positioning. A hyperbolic tangent function is applied to $\mathcal{L}_1$ to confine its range to (-1, 1), preventing the need for dynamic rescaling and enabling standardized regularization across diverse metrics.

The purpose of the total variation regularization term R_{TV} is to penalize variations between model-estimated inter-element Euclidean distances d and multiples of the element pitch p. This square difference is summed over the first $M = 10$ lags. Here, the lag index m enables the calculation of the distance between an element and another element m neighbors away. The number of lags was empirically chosen and can be increased to de-emphasize aperture locality. The expression for R_{TV} is

$$R_{\text{TV}}(\vec{\boldsymbol{r}}_a) = \sum_{m=1}^{M} \sum_{n=1}^{N-m} \left(d_{n,m} - m \cdot p\right)^2 , \tag{15}$$

where $d_{n,m}$ is the distance between element positions:

$$d_{n,m} = \left\| \boldsymbol{r}_{n+m} - \boldsymbol{r}_n \right\|_2 .$$

The purpose of the shape differential regularization term $R_{\text{d}z/\text{d}x}$ is to penalize the first derivative (i.e., the rate of change) of the z components with respect to the x components of the element positions in the aperture, effectively penalizing the local curvature of the estimated array shape. This is done by calculating the mean squared error difference in surface normals of each element over the first $M = 10$ lags using the index m. The same number of lags is used in both R_{TV} and $R_{\text{d}z/\text{d}x}$ to promote the same aperture locality. The expression for $R_{\text{d}z/\text{d}x}$ is

$$R_{\text{d}z/\text{d}x}(\vec{\boldsymbol{r}}_a) = \sum_{m=1}^{M} \frac{1}{N-m} \sum_{n=1}^{N-m} \left(\theta_{n+m} - \theta_n\right)^2 , \tag{16}$$

where θ_n is the finite difference in neighboring element surface normals given by Eq. 3.

III. Methods

A. Simulation

Nine unique transducer configurations were simulated to evaluate the performance of the proposed method: two convex, two concave, four sinusoidal, and one linear array. The transducer configurations are shown in Table II. For arrays 1-4 (convex and concave), the defined radius of curvature was used along with sine and cosine functions to parameterize the curved surface of the array. For arrays 5-8 (sinusoidal), the sinusoid frequency was determined such that the array size fits within one period. The sinusoid was laterally shifted using spatial-phase offsets to manipulate its concavity. Array 9 was initialized as a line in the lateral dimension. For each array geometry, the arc length of the parametric curve was defined as $\ell = (N-1) \cdot p$, where N is the number of elements and p is the pitch of the array. This variable arc length accommodates both simulation and experimental array geometries (e.g., the linear L12-3v and curvilinear C5-2) with varying numbers of elements and pitch such that a sinusoid, for example, completes a full cycle within the bounds of the specific array. The array configurations utilized a center frequency of 7.5 MHz, 128 elements, and a pitch of 200.0 µm.

TABLE II
Simulated Array Geometries

Array	Details
1	Convex, Radius: +50 mm
2	Concave, Radius: -50 mm
3	Convex, Radius: +10 mm
4	Concave, Radius: -10 mm
5	Sinusoid (1 cycle)
6	Sinusoid (1 cycle), $+\pi$ phase shift
7	Sinusoid (1 cycle), $+\pi/2$ phase shift
8	Sinusoid (1 cycle), $-\pi/2$ phase shift
9	Linear

Ten in-silico phantoms were produced and are shown in Fig. 2: one point-grid target (point spacing 2.0 mm), one speckle phantom, one multi-target phantom with various echoic inclusions and point targets, one Stanford logo-based phantom, and six ImageNet-based speckle phantoms [28].

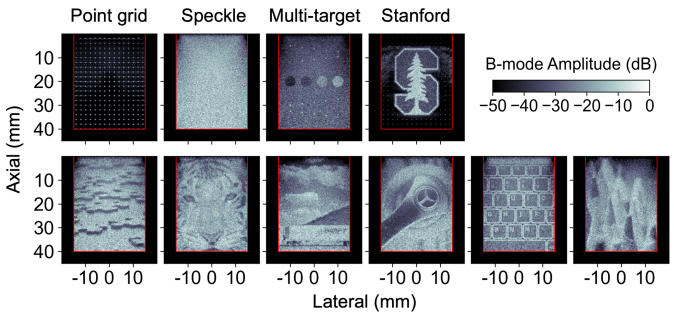


Fig. 2. Resulting B-mode images for custom (first row) point grid, speckle, multi-target, Stanford logo-based, and (second row) ImageNet-based Field II target simulations. The red rectangular region represents the areas used to compute the quality metrics and B-mode reconstructions to prevent sampling outside Field II simulation boundaries.

Field II simulations of RF channel signals were produced using the `calc_scat_multi` function with a full synthetic aperture (FSA) transmit sequence to parallelize the simulations across a CPU cluster. A sampling frequency of 160 MHz and a center frequency of 7.5 MHz were used. The ImageNet-based in silico phantoms were created using a methodology similar to Hyun et al. [29] and Brickson et al. [30]. These natural images provide a diverse set of patterns, contrasts, and heterogeneity ideal for model validation before moving to experimental phantom and in vivo data. Each in silico phantom was defined to have a sound speed of 1540 m/s and no attenuation. First, the natural images were loaded into MATLAB, converted to grayscale, normalized in the range (0,1], and interpolated onto a grid. A scatter density of 10 scatterers per resolution cell was used, where a resolution cell

was assumed to have a volume λ^3 [31]. This scatter density is approximately equivalent to 1,155 scatterers/mm^3. Field II simulations were produced for all 10 phantom layouts using each of the nine array geometries. In total, 90 simulations were generated.

After simulation, the RF data was demodulated into baseband in-phase and quadrature (IQ) components, low-pass filtered, and decimated by a factor of eight. The resulting downsampled multistatic IQ dataset $\mathbf{\Psi} \in \mathbb{C}^{N_s \cdot \mathrm{Tx} \times \mathrm{Rx}}$, array element positions $\vec{\boldsymbol{r}}_a$, downsampled sampling frequency f_s, demodulation frequency f_d (which was equal to the center frequency), and initial time t_0 values were loaded into the UltraFlex framework.

B. Additive Noise

The impact of white noise on shape estimation model performance was investigated using an additive noise model applied to simulation data. Here, white noise represents varying levels of thermal noise present in ultrasound imaging systems. Let $\mathbf{\Psi}_n(t)$ be the normalized signal representing the full synthetic receive dataset. Let $n(t)$ be the additive Gaussian white noise expressed as

$$n(t) = n_I(t) + j \cdot n_Q(t), \tag{17}$$

where $n_I(t)$ and $n_Q(t)$ are the in-phase and quadrature components, respectively, and are independent Gaussian random variables with zero means and variances σ^2 (i.e., $n_I(t), n_Q(t) \sim \mathcal{N}(0, \sigma^2)$). The noise, $n(t)$, is scaled by $\alpha = \mathbf{\Psi}_{\mathrm{rms}} \cdot 10^{(\eta/20)}$, where $\mathbf{\Psi}_{\mathrm{rms}}$ is the root-mean-square of the pulse-echo signal (which is 1 after the normalization process described in Sec. III-D), and η is the noise level in decibels (dB). The final expression for the signal with the additive noise is

$$\mathbf{\Psi}_\eta(t) = \mathbf{\Psi}_n(t) + \frac{1}{\sqrt{2}} n(t) \cdot \alpha. \tag{18}$$

While the shape estimation model configuration remained the same for each noise level, full synthetic receive data $\mathbf{\Psi}_\eta(t)$ with varying levels of noise ($\eta = [-\infty, 0, 6.0, 12.0, 18.0]$ dB) were added to the simulation data, resulting in signal-to-noise ratios (SNR) of $\mathrm{SNR}_{\mathbf{\Psi}_\eta} = [\infty, 0, -6.0, -12.0, -18.0]$ dB.

C. Experimental Data Acquisition

Experimental data was obtained from three sources using a Vantage 256 research ultrasound system (Verasonics Inc., Kirkland, WA, USA). First, RF channel data from an ATS 549 phantom (Sun Nuclear, Norfolk, VA) with a previously-calibrated ground truth sound speed of 1460 m/s [32] was acquired for this work using linear L12-3v and curvilinear C5-2v rigid transducer arrays. Second, RF channel data from three in vivo rat livers using an L12-3v transducer was obtained from a previously acquired dataset [33]. Third, in vivo human liver data was acquired using a C5-2v array under Protocol IRB-56630 and informed consent was obtained. The simplified assumption of a homogeneous sound speed distribution of 1540 m/s was made for both in vivo datasets. In total, six samples were selected from the L12-3v datasets (three phantom and three in vivo rat liver), and six samples were selected from the C5-2v datasets (three phantom and three in vivo human liver). While the ground truth array shape is known for both transducers, the model can be initialized with any array geometry to emulate a flexible array and study the impact of unknown model initialization on the subsequent shape estimation capability and convergence of the model. Table III summarizes the different acquisition configurations for each setup. In all cases, a full synthetic aperture (FSA) transmission sequence was used. The RF channel data was converted to baseband IQ data and low-pass filtered using the bandwidth f_s.

TABLE III
SUMMARY OF EXPERIMENTAL DATA ACQUISITION

Trans.	Source	c (m/s)	f_c (MHz)	f_d (MHz)	Elem.*
L12-3v	ATS pha.	1460†	6.00	6.25	128
L12-3v	Rat liver	1537‡	7.81	7.81	128
C5-2v	ATS pha.	1460†	4.00	3.91	64
C5-2v	Human liver	N/A	4.00	3.91	64

*The L12-3v and C5-2v arrays have 192 and 128 elements, respectively; only the center N elements were selected. †The ATS phantom sound speed was previously calibrated [32]. ‡The rat liver sound speed was estimated by averaging ground truth values from corresponding specimens in [33].

For both the L12-3v phantom and rat liver datasets, only data from the center 128 array elements were used to reduce the runtime and memory usage of the shape estimation model. For both the C5-2v phantom and human liver datasets, only data from the center 64 array elements were used to avoid model errors due to poor coupling of the outer elements.

D. Model Implementation

The UltraFlex framework described in Sec. II, including the differentiable image reconstruction, quality metrics, and objective function, was implemented using JAX [19]. When the IQ data $\mathbf{\Psi}(t)$ is loaded into the model, it is first normalized by the maximum IQ magnitude so that all entries have a magnitude in the range [-1,1]. Then, the data is normalized by the root mean square (RMS) magnitude. This normalization process promotes numerical floating-point stability. After this normalization process, the additive noise is introduced. For model optimization, the Adam optimizer [34] was used within the JAX-based Optax library.

1) Array Initialization: Element positions from all simulation and experimental datasets were initialized with one or more of the geometries described in Table II. The same array initialization code structure was used for Field II simulations in MATLAB and the UltraFlex framework in Python to minimize numerical differences between programming languages.

2) Reconstruction Grid Implementation: The quality metrics described previously are evaluated on patches within the imaging domain. Each *patch* is centered at a specific 2D position $\boldsymbol{r}_p$ and includes a local 3 x 3 (lateral by axial) kernel of relative offsets $\vec{\boldsymbol{\delta}}_k$ spanning the range $[-\lambda/2, +\lambda/2]$. The set of absolute positions corresponding to one patch is given by $\vec{\boldsymbol{r}}_{\mathrm{patch}} = \boldsymbol{r}_p + \vec{\boldsymbol{\delta}}_k$, where $\vec{\boldsymbol{r}}_{\mathrm{patch}} \subset \vec{\boldsymbol{r}}_{\mathrm{patches}} \in \Omega$. The full collection of patch-sampled positions across the domain is denoted by $\vec{\boldsymbol{r}}_{\mathrm{patches}}$, and is distinct from the higher-resolution imaging grid $\vec{\boldsymbol{r}}_{\mathrm{img}} \in \Omega$ used for beamformed B-

Algorithm 1 Reconstruction Grid Definitions

Define image and patch limits (mm):
if array is "C5-2v" **then**
  $\boldsymbol{x}_{\text{img\_range}} \in [-100, 100]$, $\boldsymbol{z}_{\text{img\_range}} \in [0, 140]$
  $\boldsymbol{x}_{\text{patch\_range}} \in [-100, 100]$, $\boldsymbol{z}_{\text{patch\_range}} \in [20, 72]$
else
  $\boldsymbol{x}_{\text{img\_range}} \in [-15, 15]$, $\boldsymbol{z}_{\text{img\_range}} \in [0, 40]$
  $\boldsymbol{x}_{\text{patch\_range}} \in [-15, 15]$, $\boldsymbol{z}_{\text{patch\_range}} \in [2, 42]$
end if
Compute image reconstruction grid:
$\boldsymbol{x}_{\text{img}} \leftarrow$ Linspace in $\boldsymbol{x}_{\text{img\_range}}$ with spacing $\lambda/3$
$\boldsymbol{z}_{\text{img}} \leftarrow$ Linspace in $\boldsymbol{z}_{\text{img\_range}}$ with spacing $\lambda/3$
$\vec{\boldsymbol{r}}_{\text{img}} \leftarrow$ Cartesian product of $\boldsymbol{x}_{\text{img}}$ and $\boldsymbol{z}_{\text{img}}$
Compute patch center grid:
$\boldsymbol{x}_{\text{p}} \leftarrow$ Linspace in $\boldsymbol{x}_{\text{patch\_range}}$ with 18 points
$\boldsymbol{z}_{\text{p}} \leftarrow$ Linspace in $\boldsymbol{z}_{\text{patch\_range}}$ with 21 points
$\vec{\boldsymbol{r}}_{\text{p}} \leftarrow$ Cartesian product of $\boldsymbol{x}_{\text{p}}$ and $\boldsymbol{z}_{\text{p}}$
Define local kernel offsets:
$\boldsymbol{\delta}_k \leftarrow$ Cartesian product of lateral and axial linspaces in $[-\lambda/2, \lambda/2]$
Compute full patch sampling grid:
$\vec{\boldsymbol{r}}_{\text{patches}} \leftarrow \vec{\boldsymbol{r}}_p + \vec{\boldsymbol{\delta}}_k$ (vector sum of patch centers and kernel)

mode reconstruction. Pseudo-code for initializing the imaging and patch-level grids is provided in Alg. 1. An 18 x 21 (lateral by axial) grid of patch centers $\vec{\boldsymbol{r}}_p$ was used, with spacings of 1.67 mm $\times$ 1.90 mm for regular field-of-view apertures (simulations and L12-3v) and 11.11 mm $\times$ 2.48 mm for larger apertures (C5-2v). For metrics employing synthetic subapertures in common-midpoint analysis, a subaperture size of 17 elements was used. For B-mode reconstruction, a grid spacing of $\lambda/3$ was applied.

3) Regularization: The impact of the two regularization terms was examined by a 2D parametric sweep of term weightings, where $w_2 = [10^{-4}, 10^{9}]$ and $w_3 = [10^{-9}, 10^{1}]$ in steps of powers of ten. The evaluation was performed across the speckle simulations produced using the nine array geometries to determine the ideal weightings for each regularization term of Eq. 14. The model was initialized with the linear array (Array 9) for each speckle simulation dataset.

E. Model Evaluation

1) Model Evaluation Criterion: The mean Euclidean error (MEE) $\mathcal{E}$ between the estimated element positions and ground truth element positions was used as the main evaluation criterion across all experiments and is defined as

$$\mathcal{E}^{(i)} = \frac{1}{N_a}\sum_{a=1}^{N_a}\left\|\boldsymbol{r}_a^{(i)} - \boldsymbol{r}_a^{(\text{GT})}\right\|_2 = \frac{1}{N_a}\sum_{a=1}^{N_a}\sqrt{(x_a^{(i)} - x_a^{(\text{GT})})^2 + (z_a^{(i)} - z_a^{(\text{GT})})^2} \tag{19}$$

where N_a is the total number of elements in the aperture, $\boldsymbol{r}_a^{(i)}$ represents a model-estimated element position at iteration i, and $\boldsymbol{r}_a^{(\text{GT})}$ represents the corresponding ground truth element position. MEE is used to evaluate model performance for different quality metrics and SNRs. We note that Eq. 19 is identical to the mislabeled "mean absolute error (MAE)" used in previous work and is thus directly comparable to the errors described in previous literature.

2) Model Convergence Validation: The convergence of the iterative model over $P = 1000$ iterations is examined using the following stability error criterion: During the last $K = 100$ iterations, the mean displacement of the element positions must be less than or equal to 100 µm. The mean magnitude of the displacement of element positions over the last K iterations and all N_a elements is

$$\bar{\Delta} = \frac{1}{PN_a}\sum_{i=P-K}^{P-1}\sum_{a=1}^{N_a}\left\|\boldsymbol{r}_a^{(i+1)} - \boldsymbol{r}_a^{(i)}\right\|_2 \tag{20}$$

given iteration i and element index a. Notably, this convergence validation is independent of the true element positions.

3) Computational Resource Usage: Model development was explored on an NVIDIA RTX 3090 with 24 GB of VRAM (NVIDIA, Santa Clara, CA) and then deployed on a SLURM cluster comprised of NVIDIA RTX A6000 GPUs with 48 GB of VRAM. Finally, to demonstrate the computational accessibility of this work, model runtime results for an NVIDIA RTX 3060 with 12 GB of VRAM are shown. In all cases, JAX was used with `jaxlib` built for CUDA 12.2.

IV. Results

A. Simulation

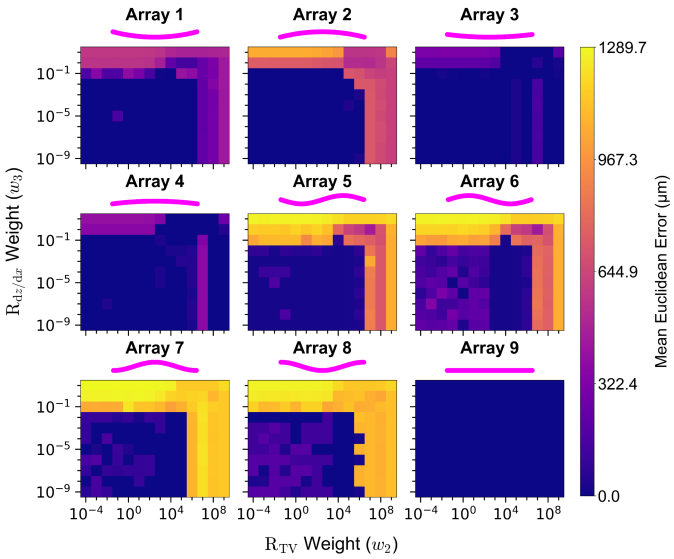


Fig. 3. Comparison of regularization method weightings for each of the nine array geometries. The shape estimation model was initialized with the linear array (Array 9) for a speckle target (top row, second column, Fig. 2) while using the filtered common-midpoint phase-error model. No additive noise was applied (i.e., $\text{SNR}_{\boldsymbol{\Psi}_{\boldsymbol{\eta}}} = \infty$).

1) Regularization: The parametric sweep results of regularization weights are visualized in Fig. 3. For the particular model configuration described in the caption of Fig. 3, shape-differential regularization ($R_{\text{dz/dx}}$) had no impact on MEE results for $w_3 < 10^{-1}$. At this weighting and above ($w_3 \geq 10^{-1}$), shape differential regularization adversely affected MEE results. Additionally, TV regularization (R_{TV}) did not affect the MEE results for $w_2 < 10^{3}$, followed by an improvement in estimation on weighting range $w_2 = [10^{3}, 10^{6})$. For $w_2 \geq 10^{6}$, MEE results for Arrays 5-8 (i.e.,

sinusoidal geometries) were adversely affected, MEE results for Arrays 1-4 (convex and concave geometries were either adversely effected or exhibited little change, and MEE results for Array 9 (linear array) did not change. Based on these results, the remaining experiments presented in this paper utilize a model configuration with TV regularization only and a term weighting of $w_2 = 10^4$.

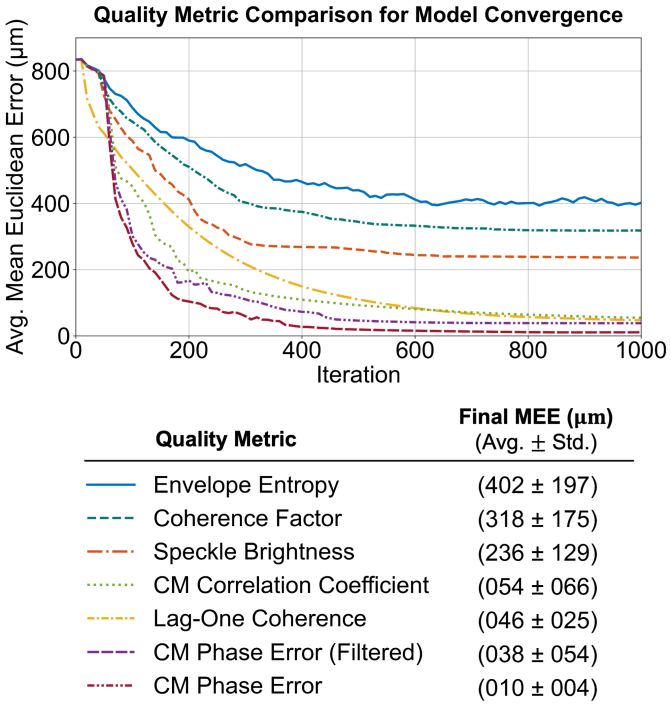


Fig. 4. Average MEE vs. model iteration results for various metric-based objective functions. For each quality metric, the MEE results are averaged across the nine array configurations and the four simulation targets (i.e., the top row of Fig. 2). The model was initialized with the linear array geometry, and a regularization weighting of $w_2 = 10^4$ was used in all cases. No additive noise was applied.

2) Metric Comparison: In Fig. 4, average MEE across the simulated data for various metric-based objectives are plotted against model iteration. The envelope entropy metric results in the worst shape estimation model performance, characterized by the greatest final average MEE. Coherence factor and speckle brightness exhibit similar model performance with final average MEEs of 318 µm and 236 µm, respectively. The correlation coefficient, lag-one coherence, and phase-error metric-based models result in the best shape estimation performance, with final average MEEs less than or equal to 54 µm.

An example of B-mode reconstructions at various iterations using the filtered phase-error metric-based model is shown in Fig. 5. Between iterations 70 and 130, image structures come into focus and shape estimation becomes more accurate. Between iterations 230 and 1000, smaller improvements in shape estimation occur. Improvements in image focusing and quality are harder to identify visually. Very small improvements may be observed between iteration 230 and higher, although almost no visually detectable changes are observed after iteration 250 despite the reduction in MEE. Additionally, MEE decreases from 62.8 µm to 19.3 µm.

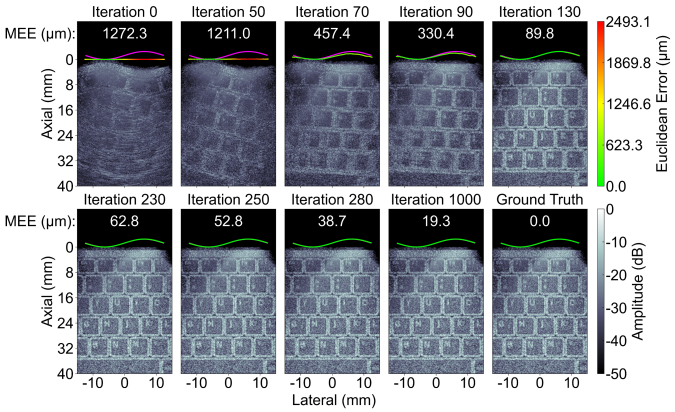


Fig. 5. Comparison of focusing at various iterations. A pink curve represents the ground-truth array shape, and the overlaid curve represents the model-estimated array shape with a variable color map from green to red representing lower and higher element-wise Euclidean errors, respectively. The array-wise MEE, in µm, is displayed above each array.

Fig. 6 exhibits B-mode focusing across different quality metrics for selected simulations. Compared with the quantitative results averaged across all nine array geometries shown in Fig. 4, the results shown in Fig. 6 correspond to Array 5 only. The envelope entropy metric-based model fails to estimate the array shape for all of the phantoms. The correlation coefficient and phase-error metric-based models achieved the best focusing of the point grid phantom target, evident in the sharpness of the point-spread function on and near the point targets.

TABLE IV
MODEL VALIDATION RESULTS

Quality Metric	Failed Sim. Cases (Fig. 7)		Failed Exp. Cases (Fig. 9)	
	No Reg.	Reg.	ATS	Liver
Envelope Entropy	146	0	6	25
Coherence Factor	7	0	0	0
Speckle Brightness	9	0	0	0
CMCC	2	0	1	0
Lag-One Coherence	58	3	0	0
CMPE (Filtered)	6	0	0	0
CMPE	2	1	0	0
Valid Cases	1660	1886	371	353
Total Cases	1890	1890	378	378

3) Additive Noise: Fig. 7 exhibits the impact of additive white noise on the final-iteration MEE. Before plotting, the model results were validated by the procedure described in Sec. III-E.2, and only the model results that pass the validation criterion are included in Fig. 7. Table IV summarizes the simulation validation process results in the first two columns.

Starting with the case of no model regularization, Fig. 7a shows boxplots of the MEE for the final iteration, arranged in descending order of median value from left to right. The inter-quartile range (IQR) for envelope entropy is small in this case because many model results did not meet the validation criterion. Notably, the boxplot includes model results above the initial average MEE. This means that in some cases, the shape estimation model produced MEEs that were further from the

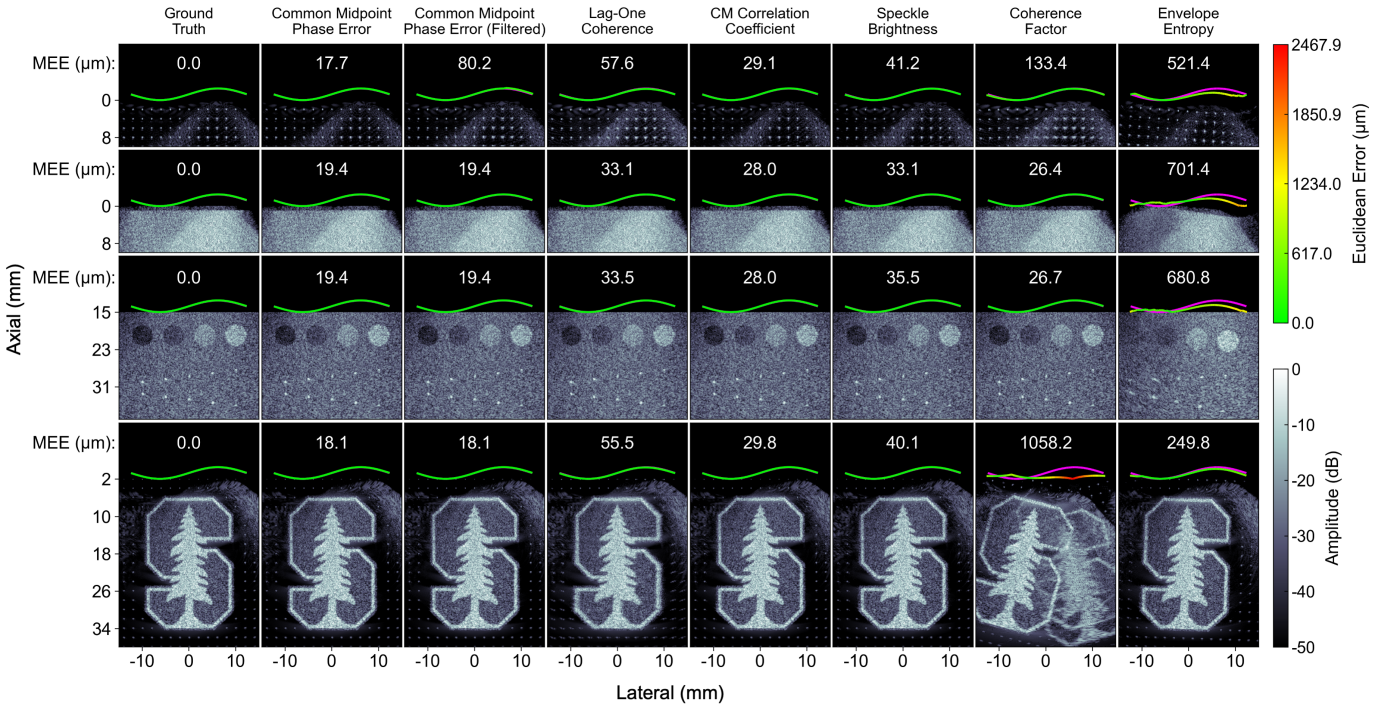


Fig. 6. B-mode images based on model convergence using different image quality metrics. Each column represents a different image quality metric used in the objective function of the iterative model. The columns are organized from left to right based on the ascending order of the final average MEEs presented in Fig. 4. From top to bottom, a point grid, speckle, multi-target, and Stanford logo-based phantom are visualized in each row. A pink curve represents the ground-truth array shape, and the overlaid curve represents the model-estimated array shape with a variable color map from green to red representing lower and higher element-wise Euclidean errors, respectively. The array MEE, in μm, is displayed above each array.

ground truth than if no shape estimation model had been used. In Fig. 7b, a partial boxplot is shown for the median and IQR of each metric group. The same metric order is maintained for all SNR groupings established in Fig. 7a. As the additive noise level increases, all median MEEs monotonically increase with the exception of envelope entropy and lag-one coherence. The lag-one coherence metric-based model produces no valid results at $\text{SNR}_{\mathbf{\Psi}_\eta} = -18.0$ dB.

Fig. 7c shows the same information as Fig. 7a but with model regularization ($w_2 = 10^4$). With regularization, the lag-one coherence, correlation coefficient, and phase-error metrics have a much smaller IQR compared to the other metrics. The coherence factor results have the largest IQR despite having one of the lowest median values.

B. Experimental

The proposed shape estimation models were validated using experimental phantom and in vivo datasets. For Fig. 8a-b, the first three columns show various shape initialization approaches for data from the calibrated ATS 549 phantom. The second three columns show various shape initialization approaches for in vivo liver data from a rat (L12-3v only) and a human (C5-2v only). In each case, the filtered phase-error metric-based model was used. The model converges to the "correct" image in all cases despite sound-speed assumption errors and different initialization configurations. However, for the ground-truth-initialized linear and curvilinear examples, the post-convergence shapes contain non-zero MEEs.

In all results shown in Fig. 8, the qualitative focusing of the resulting B-modes improves. For the L12-3v phantom data results, speckle becomes brighter and all cyst targets become visible. Furthermore, for the L12-3v in vivo rat liver data, the muscle fascia between the depths of 1.0 and 5.0 mm straightens and becomes brighter, the specular tissue at x = -5.0 mm, z = 16.0 mm becomes brighter, and features below 20.0 mm increase in visual clarity.

Fig. 9 shows a quantitative comparison of experimental results from different metric-driven models. Again, before plotting, the model results were validated by the procedure described in Sec. III-E.2, and only the model results that pass the validation criterion are shown in Fig. 9. Table IV summarizes the experimental validation process results in the last two columns.

In the idealized homogeneous phantom results, most of the metric-driven models (except speckle brightness and envelope entropy) have similar final MEEs, which are larger than the noise-free simulation results. In vivo results yielded good performance with increased error compared to the simulation and phantom results. The speckle brightness metric-based model had the worst performance, representing a departure from the metric comparison results from the simulation and additive noise studies presented earlier.

C. Computational Resource Usage

Table IV-C provides a runtime comparison (1000 iterations) between the different quality metric-based models and various NVIDIA RTX GPUs. Each entry in the table includes the

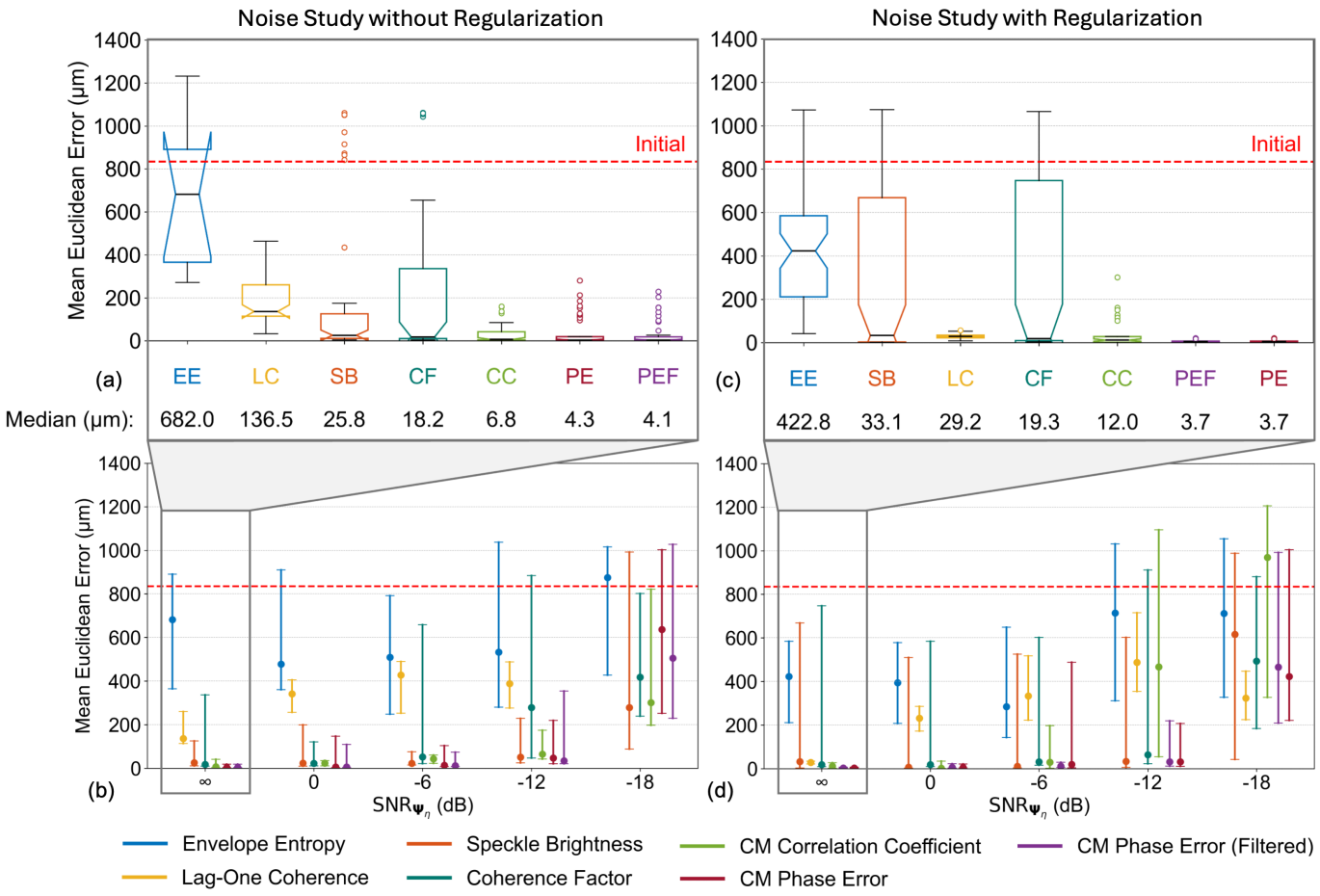


Fig. 7. Quantitative comparison of final-iteration shape estimation model performance across different image quality metrics and $\text{SNR}_{\Psi_\eta} = [\infty, \mathbf{0}, \mathbf{-6.0}, \mathbf{-12.0}, \mathbf{-18.0}]$ dB. Each boxplot (a,c) and partial boxplot (b,d) contains all validated results averaged across the nine array geometries and the six ImageNet-based speckle phantoms. The initial average MEE when using the linear array (Array 9) for model initialization is shown as a dotted red line. (a) A boxplot of model results for each quality metric is shown without model regularization, representing the first group in the (b) partial boxplot (median and IRQ) comparison of different SNR_{Ψ_η} values. (c) A boxplot of model results for each quality metric is shown with model regularization ($\boldsymbol{w_2} = \mathbf{10^4}$), representing the first group in the (d) partial boxplot comparison of various SNR_{Ψ_η} values.

TABLE V
ULTRAFLEX MODEL RUNTIME

Quality Metric	3090 (s)	A6000 (s)	3060 (s)
Envelope Entropy	147 ± 5	232 ± 19	316 ± 11
Coherence Factor	152 ± 5	223 ± 23	331 ± 23
Speckle Brightness	146 ± 4	231 ± 19	316 ± 12
Correlation Coefficient	155 ± 6	231 ± 20	344 ± 34
Lag-One Coherence	156 ± 5	227 ± 21	342 ± 29
Phase Error (Filtered)	156 ± 7	231 ± 20	346 ± 34
Phase Error	156 ± 6	231 ± 19	345 ± 34

average runtime of one simulation, one L12-3v, and one C5-2v result.

V. DISCUSSION

The UltraFlex framework was evaluated using multiple quality metrics in simulation, phantom, and in vivo data. Recent flexible-array shape calibration methods in the literature achieve an average MEE in position as low as 40 μm [18] for noise-free simulation data and as high as 1,100 μm [16] for in vivo data. With the proposed method, a median MEE as low as 3.7 μm was achieved for noise-free simulations. For intuitive understanding of these errors, visual inspection of the images in Fig. 5 and 6 shows that a 40 μm MEE and below does not appear to produce visibly detectable changes in image quality. For example, in Fig. 6, the point targets and the borders of the "S" in the Stanford logo appear nearly identical to the ground truth for the speckle brightness images, which have MEEs of 41.2 μm and 40.1 μm for the point target and Stanford logo images, respectively. Visibly apparent degradations begin to appear at 55.5 μm MEE in Fig. 6 (e.g., the vertical borders of the Stanford logo in the lag-one coherence image) and at 62.8 μm MEE in Fig. 5.

Note that the image quality from these errors is under the assumption of a correct and constant speed of sound. In Fig. 8, high-quality images are still produced despite the larger MEEs. In Fig. 8a, the final shape estimations have a small symmetric convex curvature bias. However, in the case of the L12-3v in vivo rat liver data, the final shape estimations have a small asymmetric convex curvature bias. These errors are likely the result of an incorrect sound speed used to compute the time delays in the model. For the in vivo cases, the sound speed error is heterogeneous and has a mean positive

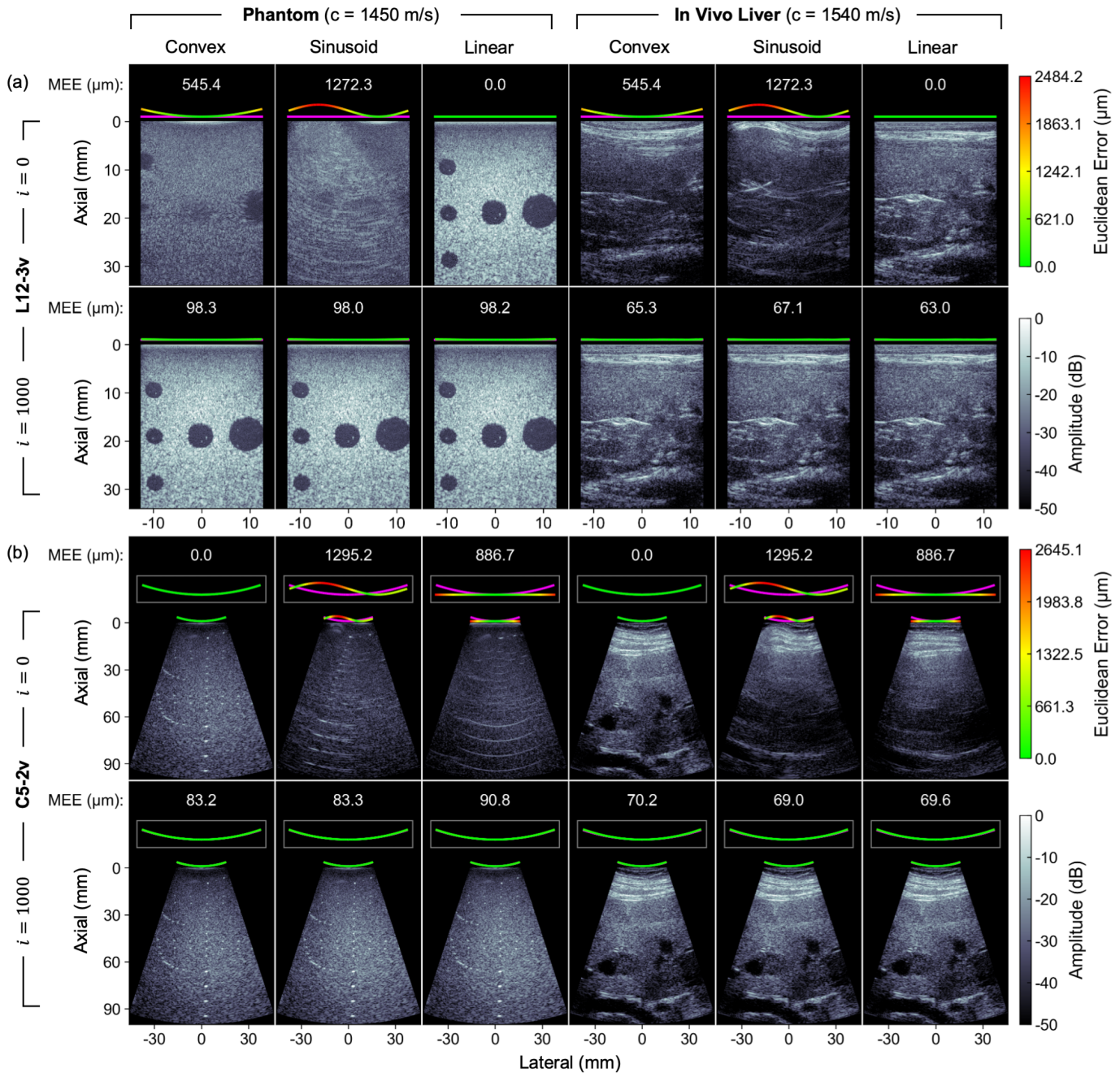


Fig. 8. B-mode images using the UltraFlex framework and filtered phase-error metric for (a) L12-3v and (b) C5-2v transducer data. Three array shape configurations are used to initialize the shape estimation model for both phantom and in vivo data: Array 1 (curvilinear; first column), Array 5 (sinusoidal; second column), and Array 9 (linear; third column). In both panels (a) and (b), the top row depicts the model initialization and resulting B-mode, whereas the bottom row depicts the final-iteration shape estimation and resulting B-mode. A pink curve represents the ground-truth array shape, and the overlaid curve represents the model-estimated array shape with a variable color map from green to red representing lower and higher element-wise Euclidean errors, respectively. The array MEE, in μm, is displayed above each array.

bias throughout the domain, resulting in a nonuniform convex curvature bias. Because both sound speed errors and element position errors result in phase error during reconstruction, the shape estimation model incorporates the sound speed error into the shape. This error is similar to a near-field phase-screen model commonly used for aberration correction. A similar deviation from the ground truth can be observed in the C5-2v phantom images of Fig. 8b. In addition to sound speed error, experimental data has noise, so the error is a combination of both sound speed error and SNR. For in vivo cases, there is also potentially diffuse reverberation noise that will contribute to increased error. Even though the in vivo shape estimations show relatively large errors compared to simulation results, the image quality remains high because the model is compensating for both the shape and the effective phase screen error. Thus, the MEE shown in Fig. 8 cannot be directly compared to the MEE of the simulations (Fig. 6) because some of the MEE in Fig. 8 is due to corrections for sound speed errors appearing

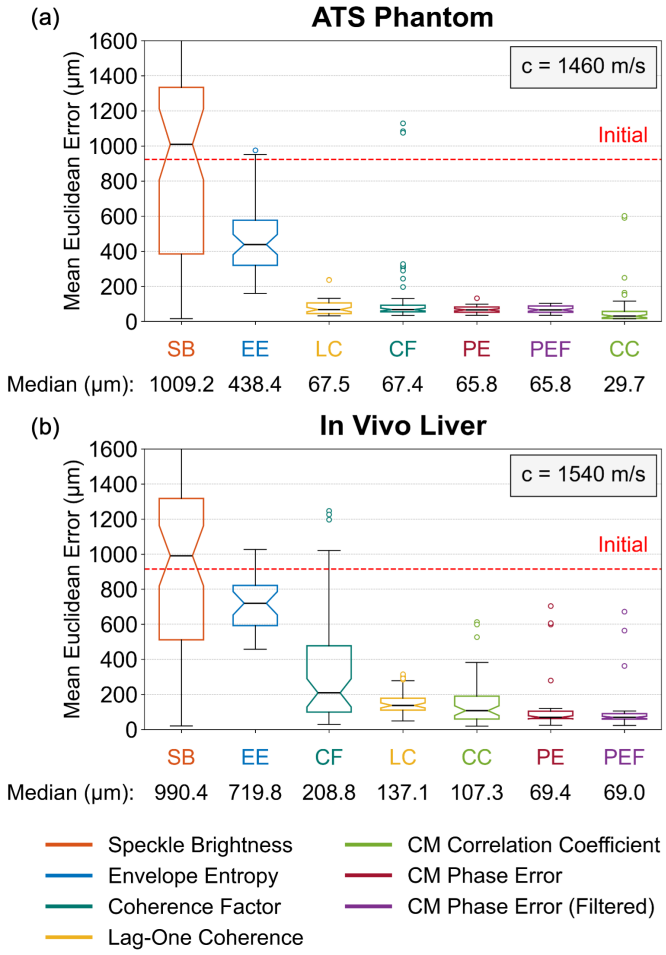


Fig. 9. Quantitative comparison of model convergence across different image quality metrics. Results are grouped by (a) ATS phantom and (b) in vivo liver data. For each grouping, three L12-3v datasets and three C5-2v datasets are utilized. The model is initialized with all nine array geometries using the indicated sound speed of each grouping. Thus, each boxplot consists of 54 results (depending on the criterion validation process).

as position errors. However, in many applications of flexible arrays, the higher-quality image may be preferable to knowing the exact position of the array elements.

The performance of the speckle brightness-based model on experimental data is significantly worse compared to its performance in the simulation and additive noise studies. This is likely due to the metric reductions in Table I being empirically tuned for simulation datasets. Thus, the tuning for speckle brightness may not be best suited for phantom and in vivo datasets because speckle brightness is inherently sensitive to amplitude distribution variations. This is a fundamental limitation of speckle brightness and entropy: optimization objective metrics with fewer local minimums or maximums will lead to better final performance than speckle brightness and envelope entropy because of their lack of sensitivity to the sampling of the amplitude distribution.

The overall performance of envelope entropy presented across all experiments in this work was substantially worse than previous reports in the literature [7], [18]. This is likely due to differences in the implementation of envelope entropy and region-of-interest (ROI) windowing. In Noda et al. [18], model evaluation was performed on very small aperture sizes of only 20 elements, ranging between 4 and 16 mm in length, which is a favorable condition for good convergence because the initialization errors are small. In this work, the larger aperture sizes inherently introduced larger initialization errors, leading to envelope entropy having the worst convergence performance in terms of the number of invalid results. Furthermore, this work used a fixed grid to compute the quality metric at image domain positions $\vec{r}_{\text{patches}}$. The spacing between the coordinates on this grid was greater than or equal to 2 mm. Therefore, localized entropy variations were not considered. For envelope entropy results presented in Fig. 7c and Fig. 7d, some of the final average MEEs were greater than the initial average MEE. These images appear to have better qualitative or visual focusing than expected, but this is usually exhibited as rotated image features that appear in focus (e.g., the coherence factor result for the Stanford logo shown in Fig. 6). This optimization state is due to the optimizer converging to a local minimum.

Different simulation targets presented unique qualitative focusing results across the various quality metrics. The point target grid sometimes had the best and the worst reconstruction compared to the other phantom targets in Fig. 6. The periodic structure of the point target grid and points in the Stanford logo presented a greater number of local minima in the objective function, thereby hampering performance. This is also potentially why we see a difference between the *average* MEE results shown in Fig. 4 and the *median* MEE results shown in Fig. 7a and Fig. 7c.

The correlation coefficient and phase-error metrics (regular and filtered) were the only results that had their average MEE vs. iteration curves intersect all other average MEE vs. iteration curves (roughly iteration 50 in Fig. 4). This suggests that in a stepwise comparison, the correlation coefficient and phase-error-based objectives could more efficiently optimize the array shape on the iteration range [50, 150] than the other objectives, despite having the same or lower learning rate. While lag-one coherence was the most efficient on the iteration range [0, 50], this is likely due to the higher learning rate used in the lag-one coherence-based model, suggesting an opportunity to further adjust learning rates to improve convergence.

Rigid-array transducers were chosen to provide a known ground truth for performance evaluation, and while the array is not flexible, a flexible array is emulated by initializing the model with an erroneous shape (e.g. curvilinear array data is examined with a sinusoidal array shape initialization). Data from the C5-2v array was limited to the central 64 elements due to element coupling between a curvilinear rigid array and a linear phantom surface; it was not due to model failure with larger arrays.

Regarding the computational resource usage of the UltraFlex framework, each model currently takes less than 2.6 minutes to achieve convergence, and a significant amount of this time is for B-mode reconstruction. Three potential approaches can significantly improve runtime. The first approach is to apply one or more of the following: reduce the number of

iterations, further tune the learning rate, or modify the reduction expressions presented in Table I. The second approach is to investigate alternative auto-differentiation frameworks, directly implement UltraFlex in a compiled language, or acquire different GPU hardware that is faster or has more VRAM, or any combination of these. The third approach is to simply use the models for initial shape calibration as opposed to a real-time implementation; for some applications, the body surfaces do not change significantly over time.

Although we exclusively use FSA sequences to establish the proof of concept, the UltraFlex framework is compatible with any imaging sequence where the receive channel data is collected for every transmit event. In the latter case, if sufficient transmit events are used, REFoCUS [35], [36] can be used to reconstruct the multistatic (FSA) data needed to apply the differentiable framework [27].

VI. Conclusion

UltraFlex, an iterative model-based ultrasonic flexible-array shape calibration framework based on automatic differentiation, was presented. Model performance was quantitatively evaluated while examining multiple image quality metrics: envelope entropy, speckle brightness, lag-one coherence, coherence factor, common-midpoint correlation coefficient, and common-midpoint phase error. These image quality metrics were evaluated on simulated phantoms using a variety of array shapes. Experimental phantom and in vivo liver datasets were also investigated using transducers with known geometries. Speckle brightness, envelope entropy, and coherence factor enabled model convergence under many conditions. Lag-one coherence, common-midpoint correlation coefficient, and common-midpoint phase error enabled more accurate element position estimations and improved visual ultrasound image focusing quality. Results indicate that common-midpoint correlation coefficient and phase-error quality metrics were the most robust against additive white noise while achieving median MEEs of 3.7 µm for simulation, 29.7 µm for phantom, and 69.0 µm for in vivo liver data. These array shape calibration results show promise for the current and future development of experimental flexible- and wearable-ultrasonic arrays.

References

[1] C. Wang et al. Continuous monitoring of deep-tissue haemodynamics with stretchable ultrasonic phased arrays. *Nat. Biomed. Eng.*, pages 749–758, 2021.

[2] D. K. Piech, J. E. Kay, B. E. Boser, and M. M. Maharbiz. Rodent wearable ultrasound system for wireless neural recording. In *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, pages 221–225, 2017.

[3] H. Hu et al. A wearable cardiac ultrasound imager. *Nature*, pages 667–675, 2023.

[4] W. Lyu, Y. Ma, S. Chen, H. Li, P. Wang, Y. Chen, and X. Feng. Flexible ultrasonic patch for accelerating chronic wound healing. *Adv. Healthc. Mater.*, page 2100785, 2021.

[5] V. Pashaei, P. Dehghanzadeh, G. Enwia, M. Bayat, S. J. A. Majerus, and S. Mandal. Flexible body-conformal ultrasound patches for image-guided neuromodulation. *IEEE Trans. Biomed. Circuits Syst.*, pages 305–318, 2019.

[6] W. Chen et al. Flexible ultrasound transducer with embedded optical shape sensing fiber for biomedical imaging applications. *IEEE Trans. Biomed. Eng.*, pages 2841–2851, 2023.

[7] X. Huang, H. Hooshangnejad, D. China, Z. Feng, J. Lee, M. A. L. Bell, and K. Ding. Ultrasound imaging with flexible array transducer for pancreatic cancer radiation therapy. *Cancers*, page 3294, 2023.

[8] D. China, Z. Feng, H. Hooshangnejad, D. Sforza, P. Vagdargi, M. A. L. Bell, A. Uneri, A. Sisniega, and K. Ding. Flex: Flexible transducer with external tracking for ultrasound imaging with patient-specific geometry estimation. *IEEE Trans. Biomed. Eng.*, pages 1298–1307, 2024.

[9] R. J. McGough, D. Cindric, and T. V. Samulski. Shape calibration of a conformal ultrasound therapy array. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, pages 494–505, 2001.

[10] J. Elloian, J. Jadwiszczak, V. Arslan, J. D. Sherman, D. O. Kessler, and K. L. Shepard. Flexible ultrasound transceiver array for non-invasive surface-conformable imaging enabled by geometric phase correction. *Sci. Rep.*, page 16184, 2022.

[11] J. Zhang, K. Ding, and M. A. L. Bell. Impact of photoacoustic source location on flexible array curvature estimation with a maximum lag-one spatial coherence metric. In *Proc. IEEE Ultrason. Ferroelectr. Freq. Control Joint Symp. (UFFC-JS)*, pages 1–4, 2024.

[12] A. J. Hunter, B. W. Drinkwater, and P. D. Wilcox. Autofocusing ultrasonic imagery for non-destructive testing and evaluation of specimens with complicated geometries. *NDT E Int.*, pages 78–85, 2010.

[13] J. Chang, Z. Chen, Y. Huang, Y. Li, X. Zeng, and C. Lu. Flexible ultrasonic array for breast-cancer diagnosis based on a self-shape-estimation algorithm. *Ultrasonics*, page 106199, 2020.

[14] M. Ingram and J. D'hooge. Estimation of flexible ultrasound array shape using phase coherence. *IEEE Trans. Biomed. Eng.*, pages 1–11, 2024.

[15] A. Omidvar, R. Rohling, E. Cretu, M. Cresswell, and A. J. Hodgson. Shape estimation of flexible ultrasound arrays using spatial coherence: A preliminary study. *Ultrasonics*, page 107171, 2024.

[16] T. Noda, T. Azuma, Y. Ohtake, I. Sakuma, and N. Tomii. Ultrasound imaging with a flexible probe based on element array geometry estimation using deep neural network. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, pages 3232–3242, 2022.

[17] X. Huang, M. A. L. Bell, and K. Ding. Deep learning for ultrasound beamforming in flexible array transducer. *IEEE Trans. Med. Imaging*, pages 3178–3189, 2021.

[18] T. Noda, N. Tomii, K. Nakagawa, T. Azuma, and I. Sakuma. Shape estimation algorithm for ultrasound imaging by flexible array transducer. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, pages 2345–2353, 2020.

[19] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang. JAX: Composable transformations of Python+NumPy programs. `http://github.com/jax-ml/jax`, 2018. Version: 0.3.13.

[20] D. Hyun, S. V. Narayan, W. Simson, L. L. Zhuang, and J. J. Dahl. Flexible array shape estimation using differentiable beamforming. In *Proc. IEEE Int. Ultrason. Symp. (IUS)*, pages 1–4, 2023.

[21] L. Nock, G. E. Trahey, and S. W. Smith. Phase aberration correction in medical ultrasound using speckle brightness as a quality factor. *J. Acoust. Soc. Amer.*, pages 1819–1833, 1989.

[22] L. Xi, L. Guosui, and J. Ni. Autofocusing of isar images based on entropy minimization. *IEEE Trans. Aerosp. Electron. Syst.*, pages 1240–1252, 1999.

[23] R. Mallart and M. Fink. Adaptive focusing in scattering media through sound-speed inhomogeneities: The van cittert zernike approach and focusing criterion. *J. Acoust. Soc. Amer.*, pages 3721–3732, 1994.

[24] W. Long, N. Bottenus, and G. E. Trahey. Lag-one coherence as a metric for ultrasonic image quality. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, pages 1768–1780, 2018.

[25] D. Rachlin. Direct estimation of aberrating delays in pulse-echo imaging systems. *J. Acoust. Soc. Amer.*, pages 191–198, 1990.

[26] G. C. Ng, W. F. Walker, and G. E. Trahey. Improvement of signal correlation for adaptive imaging using the translating transmit aperture algorithm. In *Proc. IEEE Int. Ultrason. Symp. (IUS)*, pages 1395–1400, 1996.

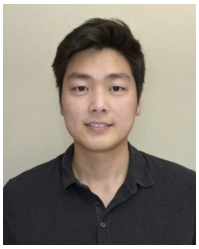
[27] W. Simson, L. Zhuang, S. J. Sanabria, N. Antil, J. J. Dahl, and D. Hyun. Differentiable beamforming for ultrasound autofocusing. In *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, pages 428–437, 2023.

[28] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 248–255, 2009.

[29] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl. Beamforming and speckle reduction using neural networks. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, pages 898–910, 2019.

[30] L. L. Brickson, D. Hyun, M. Jakovljevic, and J. J. Dahl. Reverberation noise suppression in ultrasound channel signals using a 3d fully convolutional neural network. *IEEE Trans. Med. Imaging*, pages 1184–1195, 2021.
[31] R. F. Wagner, M. F. Insana, and S. W. Smith. Fundamental correlation lengths of coherent speckle in medical ultrasonic images. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, pages 34–44, 1988.
[32] M. Jakovljevic, S. Hsieh, R. Ali, G. Chau Loo Kung, D. Hyun, and J. J. Dahl. Local speed of sound estimation in tissue using pulse-echo ultrasound: Model-based approach. *The Journal of the Acoustical Society of America*, 144(1):254–266, 2018.
[33] A. V. Telichko, R. Ali, T. Brevett, H. Wang, J. G. Vilches-Moure, S. U. Kumar, R. Paulmurugan, and J. J. Dahl. Noninvasive estimation of local speed of sound by pulse-echo ultrasound in a rat model of nonalcoholic fatty liver. *Phys. Med. Biol.*, page 015007, 2022.
[34] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
[35] N. Bottenus. Recovery of the complete data set from focused transmit beams. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, pages 30–38, 2017.
[36] R. Ali, C. D. Herickhoff, D. Hyun, J. J. Dahl, and N. Bottenus. Extending retrospective encoding for robust recovery of the multistatic data set. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, pages 943–956, 2019.

Benjamin N. Frey (Graduate Student Member, IEEE) was born on August 4, 1999 in Waconia, MN, USA. As a Schulze Innovation Scholar, he received degrees in physics (B.Sc.), computer science (B.Sc.), and business administration (B.A.) from the University of St. Thomas (Saint Paul, MN, USA) in 2022. He is currently a Ph.D. candidate in applied physics at Stanford University (Stanford, CA, USA). His research interests include computational imaging, adaptive phase correction, and wearable ultrasound.



Dongwoon Hyun (Member, IEEE) received the B.S.E. and Ph.D. degrees in biomedical engineering from Duke University (Durham, NC, USA) in 2010 and 2017. He was previously an instructor in the Department of Radiology at Stanford University (Stanford, CA, USA). He is now a senior staff engineer at Siemens Healthineers (Palo Alto, CA, USA). His research interests include adaptive beamforming and image reconstruction techniques.

Walter Simson received the B.S., M.S., and Ph.D. degrees in eletro-mechanical engineering, computational science and engineering, and computer science, respectively, from the Technical University of Munich (Munich, Germany). He was previously a Research Scientist at Stanford University (Stanford, CA, USA).

Louise Zhuang (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from the Georgia Institute of Technology (Atlanta, GA, USA) in 2020 and is currently a Ph.D. candidate in electrical engineering at Stanford University (Stanford, CA, USA). Her interests broadly involve imaging and image systems, and her current research work includes beamforming, signal processing, and image quality enhancement.

Hoda S. Hashemi (Member, IEEE) received the B.Sc. degree from Sharif University of Technology (Tehran, Iran) in 2014, the M.A.Sc. degree from Concordia University (Montreal, QC, Canada) in 2017, and the Ph.D. degree from the University of British Columbia (Vancouver, BC, Canada) in 2023, all in electrical and computer engineering. From 2021 to 2023, she was an ultrasound research intern with the Research and Innovation team at DarkVision Technologies Inc. (Vancouver, BC, Canada). She is currently a postdoctoral scholar with the Department of Radiology at Stanford University (Stanford, CA, USA). Her current research interests include ultrasound molecular imaging, elastography, and machine learning methods for medical image processing.

Martin Schneider received the M.D. degree from the University of Tübingen (Tübingen, Germany), and the Ph.D. degree in biomedical engineering from ETH Zurich (Zürich, Switzerland). He joined the Department of Radiology at Stanford University (Stanford, CA, USA) as a postdoctoral scholar. His research interests include photoacoustic imaging, ultrasound imaging, and new biomarkers.

Jeremy J. Dahl (Senior Member, IEEE) received the B.S. degree in electrical engineering from the University of Cincinnati (Cincinnati, OH, USA) in 1999, and the Ph.D. degree in biomedical engineering from Duke University (Durham, NC, USA) in 2004. He is currently an Associate Professor with the Department of Radiology at Stanford University (Stanford, CA, USA). His current interests include beamforming, aberration correction, noise suppression, sound speed imaging, and molecular imaging.