Learning-based Efficient Phase-Amplitude Modulation and Hybrid Control for MRI-guided Focused Ultrasound Treatment

Jing Dai[1]*, *Student Member, IEEE*, Bohao Zhu[1]*, Xiaomei Wang[1, 2], *Member, IEEE*, Zhiyi Jiang[1], Mengjie Wu[1], Liyuan Liang[2,3], Xiaochen Xie[4], *Senior Member, IEEE*, James Lam[1], *Fellow, IEEE*, Hing-Chiu Chang[2,3], and Ka-Wai Kwok[1], *Senior Member, IEEE*

***Abstract*—Magnetic resonance-guided focused ultrasound (MRg-FUS) has become attractive, accredited to its non-invasive nature. However, ultrasound beams focusing and steering is still challenging owing to aberrations induced by soft tissue heterogeneity. In particular for beam motion control to ensure real-time and precise tracking in the deep-seated region over abdominal organs, while considering full-wave propagation. To this end, we proposed a closed-loop hybrid control scheme and a learning-based modulation model for robot-assisted MRg-FUS treatments. By introducing a rapid phase estimator to provide an efficient (<3 ms) solution, the robust H_∞ controller enables real-time and accurate tracking (0.30 mm) without prior knowledge of heterogeneous media, even under unknown disturbances. Our model enables rapid (2.65 ms) phase-amplitude modulation and precise targeting (mean 0.35 mm, max. 0.65 mm), meeting clinical standard. Focal obliquity is significantly "aligned" to only 2.7°. Results from sensitivity analysis and transducer design also support the model's clinical feasibility and potential in widespread MRg-FUS treatments.**

***Index Terms*— Beam motion control, hybrid control, learning-based modulation, robot-assisted MRg-FUS.**

I. Introduction

FOCUSED ultrasound (FUS), owing to its advantages of being completely non-invasive and non-ionizing, has been established as promising interventions [1], e.g., treatments of uterine fibroids [2] and essential tremors [3]. Ultrasound imaging [4-6] and magnetic resonance imaging (MRI) [7, 8] have been introduced to this procedure to provide intraoperative guidance. MRI plays an increasingly important role in guidance and safety monitoring, accredited to its superior soft-tissue imaging contrast and detailed visualization of physiological changes. Magnetic resonance (MR) thermometry can offer real-time thermal maps to monitor temperature in high-intensity FUS (HIFU). MR elastography [9] can assess thermal damage by measuring tissue elasticity changes. As another elastography method, MR-acoustic radiation force imaging (MR-ARFI) [10] could detect micro-scale tissue displacements resulting from millisecond-short (1-20 ms) ultrasound pulses (**Fig. 1a**), so as to visualize focal spot without causing thermal damage [11]. It would induce negligible temperature rise, making MR-ARFI advantageous in low-intensity FUS (LIFU). Previous studies [12, 13] have reported that few MR-ARFI sequences enable rapid (~3 s) acquisition on two-dimensional (2D) displacement image, showing the potential in closed-loop feedback control. Therefore, advances in MRI are quite readily available in clinical practice for both HIFU and LIFU.

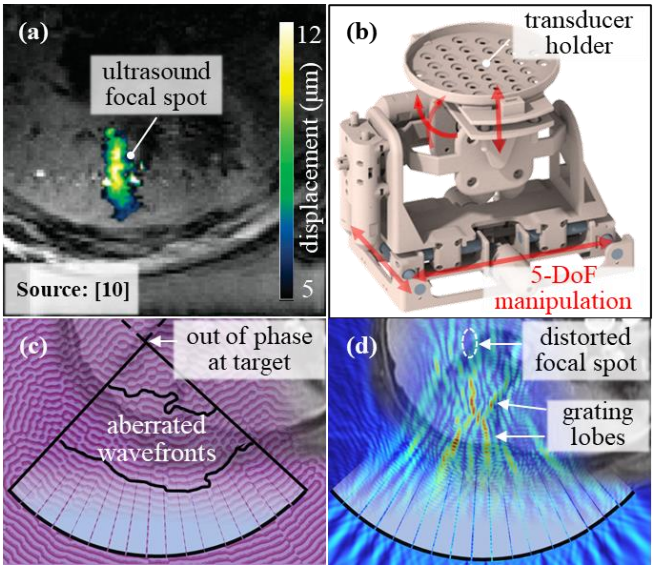

Fig. 1. (a) MR-ARFI tissue displacement map (in μm) in a sheep liver (Image source: [10]). **(b)** MRI-guided 5-DoF robotic platform providing sufficient workspace for focal spot steering. **(c)** Phase field showing phase aberration effect, where the wavefronts are out of phase due to tissue heterogeneity. **(d)** Distorted focal spot with grating lobes.

Currently, most MR-guided FUS (MRg-FUS) treatments for abdominopelvic organ diseases are in their early stage [2]. The common practice is electronic control of a fixed phased-array transducer to allow acoustic beam focusing and steering. However, it would suffer from insufficient focal-spot adjustment range, owing to the large treatment region of interest (ROI) in abdominal organs. Repeated manual re-positioning and re-calibration on the transducer is required, which can be labor-intensive and time-consuming. Integration of the robotic

Manuscript received: September 18, 2023; Revised: November 20, 2023; Accepted: November 20, 2023. This paper was recommended for publication by Editor Pietro Valdastri upon evaluation of the Associate Editor and Reviewers' comments. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62203137, in part by the Research Grants Council (RGC) of Hong Kong (17205919, 17207020, 17209021, and T42-409/18-R), in part by the Innovation and Technology Commission (ITC) (MRP/029/20X), in part by the Multi-Scale Medical Robotics Center Limited, InnoHK, and in part by the Alexander von Humboldt Foundation, Germany.

[1]J. Dai, B. Zhu, X. Wang, Z. Jiang, M. Wu, J. Lam, and K. W. Kwok are with Department of Mechanical Engineering, The University of Hong Kong, Hong Kong, China.

[2]X. Wang, L. Liang and H.C. Chang are also with Multi-scale Medical Robotics Center Limited.

[3]L. Liang and H.C. Chang are with Department of Biomedical Engineering, The Chinese University of Hong Kong, Hong Kong, China.

[4]X. Xie is with Department of Automation, Harbin Institute of Technology, Shenzhen, Shenzhen 518055, China, and Institute for Automatic Control and Complex Systems, University of Duisburg-Essen, Duisburg 47057, Germany.

K. W. Kwok (kwokkw@hku.hk) and X. Xie (xiexiaochen@hit.edu.cn) are co-corresponding authors. * indicates co-first authorships.

Digital Object Identifier (DOI): see top of this page.

system and the MRg-FUS has become a reliable solution to those challenges. It has stimulated extensive research on robotic MRg-FUS systems for abdominal and pelvic treatments, ensuring accurate (<1 mm) transducer positioning and offering sufficient focal-spot steering range. Besides, such automatic adjustment could reduce the risk of skin burn caused by long HIFU exposure. Some robot prototypes were primarily developed for pelvic FUS, such as Yiallouras *et al*. [14] and Epaminonda *et al*. [15]. Moreover, a few systems have been commercialized, e.g., Sonalleve MRg-FUS system (Philips) and ExAblate 2000 (Insightec) [2]. Instead of targeting stationary lesions in pelvic FUS, robotic systems for abdominal FUS have to track the lesion rapidly due to respiratory-induced motion, which is rare at present. Previously, we developed a 5-DoF MRI-guided robot (**Fig. 1b**) to facilitate abdominopelvic FUS [2]. Its fast focal spot tracking (>0.2 Hz) demonstrates the potential to compensate for physiological motion.

Since the robotic mechanical steering enables beams to be steered in a large coverage of the patient's abdomen, ultrasound-tissue interference would be rather complicated in making beams focusing at the target challenging. Soft tissue heterogeneity (e.g., tissue thickness) would cause phase difference along the wavefront, i.e., phase aberration (**Fig. 1c**). It inevitably results in focal shift and focus distortion [16]. Thus, electronic control of the transducer has to coordinate with robotic tele-manipulation simultaneously to result in constructive interference. Transduction planning prior to treatment has to be conducted for aberrations compensation in HIFU. To this end, several phase correction methods were proposed, such that the beams can be in phase at the target, thus allowing beam refocusing after transmitting multi-layer tissues. The hydrophone-based method [17] ensures optimal phase correction as the gold standard, but it is impractical in clinical practice due to its invasive nature [17]. Alternatively, simulation-based approaches allow precise phase-aberration correction in a non-invasive way [17]. However, phase adjustment only ensures accurate focal targeting to reduce the risk of off-target heating, which is still insufficient for effective treatment. Amplitude aberration (i.e., amplitude difference along the wavefront) is another major concern (**Fig. 1d**). It would cause unwanted side/grating lobes, and the focal spot may be oblique to the transducer's central axis, causing thermal damage to healthy tissues [18]. The solution to resolve the aberrations is precisely modulating the elements' amplitudes and phases simultaneously, so as to enable accurate targeting, lobes reduction, and focal obliquity alignment. This would improve treatment safety and efficacy. Currently, only few simulation-based methods have been explored, but they are not clinical-feasible because full-wave simulation process is time-consuming (several hours per target) [19, 20]. Moreover, they mainly resolved skull-induced aberrations for transcranial HIFU [20, 21], where brain tissues are ideally assumed to be homogenous for ease of modulation. Instead, abdominal FUS involves many layers of tissues (e.g., fat, muscle) and organs, some are inherently highly heterogeneous (e.g., liver), making transduction planning rather challenging. Besides, inevitably, numerous lesion points would have to be targeted in a large ROI over the abdominal organ, substantially prolonging treatment time [17, 19]. Thus, a time-efficient and precise aberrations modulation method is highly desired, so as to facilitate robotic MRg-FUS in widespread clinical applications.

Apart from beam refocusing at a target *point*, real-time and robust beam motion control to allow simultaneous beam refocusing and dexterous steering over a target *region* is also essential, particularly for LIFU [22]. Analytical modeling approaches can establish an inverse mapping from beam motion to the driving phase signals. The geometry-based method was developed with a homogenous medium assumption [16], but it is only applicable to homogenous or weakly heterogeneous media. Other model-based strategies, such as the ray-tracing-based method [17, 20], were explored to enable real-time open-loop control of the beam motion. The straight-line ultrasound propagation assumption is proposed to simplify the complex full-wave propagation process, thus facilitating analytical kinematics approximation without considering physical effects [23]. However, oversimplification on full-wave propagation would inevitably degrade the tracking performance. Moreover, the aforementioned methods realize point-to-point tracking in an open-loop manner, suffering from unknown disturbances, e.g., patient positioning and transducer calibration errors. Unpredictable slight physiological motion of the patient would also promptly deteriorate the analytical model.

Recent machine learning (ML) has been successfully applied in many scenarios (e.g., soft robot control [24, 25], shape sensing [26]), showing the potential to offer efficient and accurate solutions in full-wave propagation problems. In this paper, we propose a learning-based modulation model and hybrid control framework for both HIFU and LIFU. It is the *first* attempt to resolve inverse problems of full-wave aberrations correction and beam motion control using ML techniques. Our key work contributions are listed below:

1) Development of a learning-based, mesh-free, and safe compensation scheme for soft-tissue-induced aberrations for HIFU applications, which allows rapid full-wave phase-amplitude modulation for accurate beam refocusing and restoration *simultaneously*.
2) Design of a hybrid control framework which combines fast phase estimator and closed-loop robust H_∞ controller for LIFU, enabling real-time, robust tracking over a large region without requiring prior knowledge of heterogeneous media, even under unknown disturbances.
3) Proof-of-concept of the resultant phased-array design, model sensitivity, computational efficiency, phase-amplitude modulated targeting, tracking robustness, and focal obliquity alignment, in support of its clinical feasibility and potential for both HIFU and LIFU.

II. Materials and Methods

This section presents our phase-amplitude modulation method and hybrid control framework, by design for treatments over major abdominopelvic organs, such as liver, spleen, and kidney. Implementation details, such as data collection, model training and prediction are also included.

A. Framework Overview

Two major modules are taken into account, i.e., a hybrid control policy (**Fig. 2a**) and a phase-amplitude modulation model (**Fig. 2b**). The framework supports two types of positional commands. The targets can be pinpointed by an operator for HIFU ablation, and a trajectory can be predefined for target tracking in LIFU. In this proof-of-concept study, the simulator (*Section D*) can visualize the beam spot and detect the focus position, thus closing the control loop for beam motion

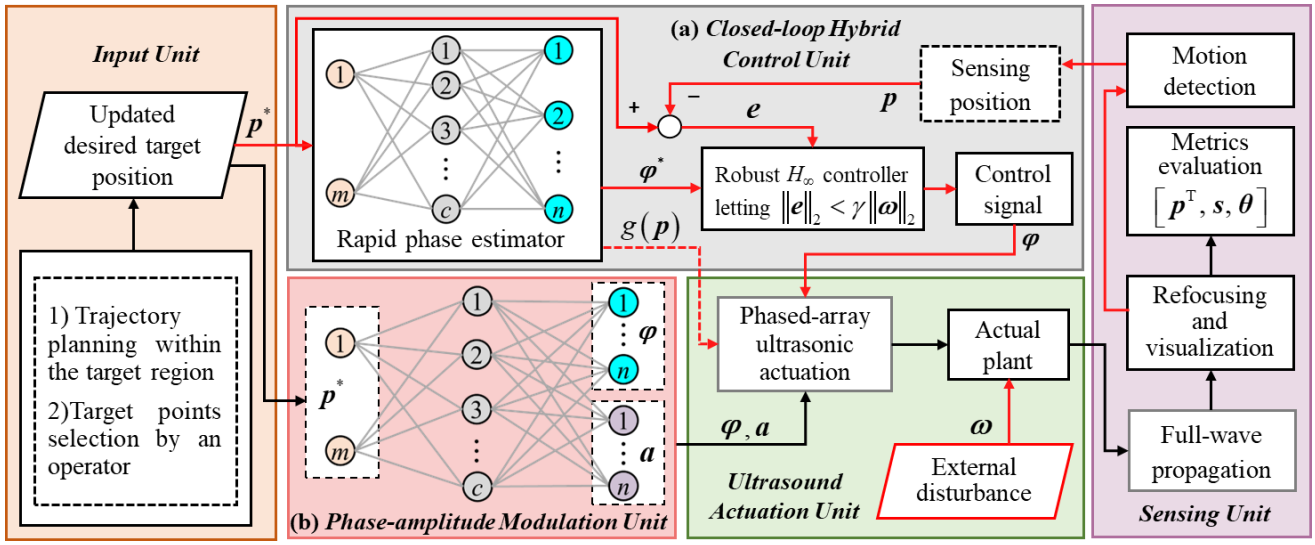

Fig. 2. Overview of the framework. **(a)** Closed-loop hybrid control scheme. The phase signal is generated by the feedback scheme and an ELM-based rapid phase estimator. **(b)** Simultaneous phase and amplitude aberrations correction. The target position acts as input of the ELM network, which is mapped to the desired phase and amplitude patterns of each element.

control. In the real world, MR-ARFI can be a reliable substitute for particular applications [12, 13].

M1) **Hybrid Control Policy:** This control scheme is specific for LIFU. The prerequisite is that the focal spot has to track consecutive targets in real time over a wide ROI (>Ø10 mm). Equal power distribution is usually applied to the elements, as amplitude aberration can be ignored owing to negligible temperature increase. This control policy circumvents the kinematics mapping by combining a phase estimator and a model-based controller (i.e., robust H_∞ controller). The ML technique is introduced to "train" the mapping between focus position and modulated phase, without prior knowledge of heterogeneous tissues. Such a learning-based estimator should offer phase actuation rapidly (>100 Hz) to the H_∞ controller, so as to guarantee fast convergence for real-time control. Accurate (mean error <5°) phase prediction is also necessary to ensure robust control, even under unknown disturbances.

M2) **Phase-Amplitude Modulation:** This model is specifically designed for HIFU, which could be pretrained for either thermal or mechanical ablation. Both phase and amplitude aberrations can be corrected simultaneously. To validate the feasibility in widespread applications, beam refocusing quality has to be tested over a large (>Ø10 mm) and deep-located (>75 mm depth) ROI in a solid organ (e.g., liver). As lesion points need to be treated, modulation should be rapid (e.g., <10 s per target) to substantially decrease treatment time. Regarding phase modulation, it is expected to enable targeting error smaller than 1 mm [27] to alleviate off-target heating. It can also be used to guide compact design of the transducer. Since tradeoffs between transducer size and manufacturing cost, fabrication effort, and assembly complexity are inevitable. Rapid modulation is expected to accelerate transducer parameters optimization, e.g., aperture size. Amplitude modulation aims to further refine the focusing quality and improve HIFU safety. By optimizing the elements' power distribution, the grating/side lobes should be reduced, and focal spot needs to be "aligned" towards the transducer's central axis.

B. Learning-based Phase-Amplitude Modulation

Inverse mapping from transducer actuation space to task space needs to be defined. The actuation space is denoted as phase pattern $\boldsymbol{\varphi} = [\varphi_1, \varphi_2, \cdots, \varphi_n]^{\mathrm{T}} \in \mathbb{R}^n$ and amplitude pattern $\boldsymbol{a} = [a_1, a_2, \cdots, a_n]^{\mathrm{T}} \in \mathbb{R}^n$ of the array, where n represents the number of elements. The task space is defined as $[\boldsymbol{p}^{\mathrm{T}}, s, \boldsymbol{\theta}]^{\mathrm{T}}$, where $\boldsymbol{p} \in \mathbb{R}^m$, $s \in \mathbb{R}$, and $\boldsymbol{\theta} \in \mathbb{R}$ are, respectively, the focal point position ($m = 2$), focus area, and focal obliquity.

Here, an extreme learning machine (ELM) [28] network is employed to establish the inverse mapping, attributed to its rapid weight initialization and model convergence. The network's neuron number is expressed by c, which can be adjusted for robust model training (c = 1500 in this study). Given a target, both the phase and amplitude patterns can be directly predicted through the learned mapping, thus allowing simultaneous phase and amplitude aberrations correction.

Training: To determine the inverse mapping, we predefine N_0 samples that cover the ROI for model training. The target position, corresponding phase, and amplitude patterns are

$$\boldsymbol{\Psi} = \left[\boldsymbol{\varphi}(1), \boldsymbol{\varphi}(2), \cdots \boldsymbol{\varphi}(N_0)\right] \in \mathbb{R}^{n \times N_0},$$

$$\boldsymbol{A} = \left[\boldsymbol{a}(1), \boldsymbol{a}(2), \cdots \boldsymbol{a}(N_0)\right] \in \mathbb{R}^{n \times N_0},$$

and

$$\boldsymbol{P} = \left[\boldsymbol{p}(1), \boldsymbol{p}(2), \cdots \boldsymbol{p}(N_0)\right] \in \mathbb{R}^{2 \times N_0},$$

respectively. Utilizing the ELM-based model, with $\boldsymbol{P}$ as input and $[\boldsymbol{\Psi}, \boldsymbol{A}]$ as output, the model can be represented as

$$\left[\boldsymbol{\varphi}(i), \boldsymbol{a}(i)\right] = h\left(\boldsymbol{p}(i)\right), \quad i = 1, 2, \ldots N_0 . \tag{1}$$

Prediction: Given a query input set $\boldsymbol{p}^*(k)$ at the k^{th} target point, the phase and/or amplitude pattern can be predicted by

$$\left[\boldsymbol{\varphi}_w(k), \boldsymbol{a}_w(k)\right] = h\left(\boldsymbol{p}^*(k)\right), \quad k = 1, 2, \ldots . \tag{2}$$

The superscript "*" denotes the desired variable specified by users or other reference input.

C. Hybrid Control for Real-time Tracking

To ensure reliable and real-time tracking within a large ROI inside the abdominal organ, it is essential to detect the focal point's position as feedback. This detection allows for the formation of a closed-loop system, which can effectively counteract disturbances and enhance robustness. Based on the

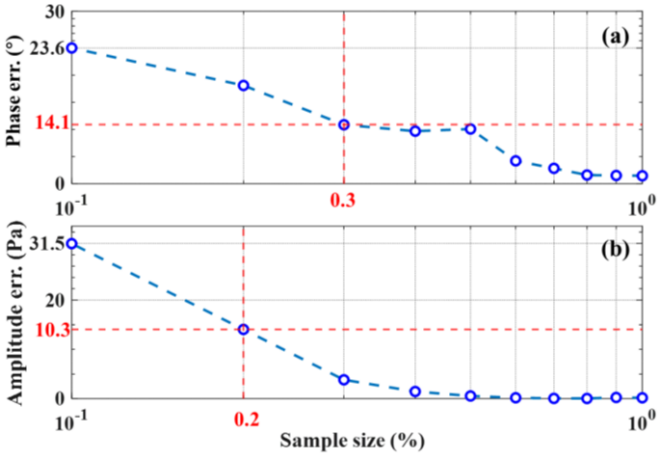

Fig. 3. Sensitivity analysis of **(a)** phase pattern and **(b)** amplitude pattern. The training sample number is varied from 10% to 100% of the total samples.

proposed goal, robust H_∞ control [24, 25] is applied to mitigate disturbances and enhance the system's robustness. The discrete-time nonlinear system is given as follows:

$$\boldsymbol{p}(k+1) = f\left(\boldsymbol{p}(k), \boldsymbol{\varphi}(k), \boldsymbol{\omega}(k)\right), \tag{3}$$

where $\boldsymbol{p}(k)$, $\boldsymbol{\varphi}(k)$, and $\boldsymbol{\omega}(k)$ are the focal point position, control input of the phase, and the unknown external disturbance, respectively. Without loss of generality, we assume that the disturbance is bounded:

$$\|\boldsymbol{\omega}\|_2 = \sqrt{\sum_{i=0}^{\infty} |\boldsymbol{\omega}(k)|^2} < \infty . \tag{4}$$

A learning-based phase estimator can be pretrained to offer information of the nonlinear vector function $f(\cdot)$ in real time. The function $f(\cdot)$ is derived from the estimator represented as $\boldsymbol{\varphi}(k) = g(\boldsymbol{p}(k))$. Since the objective is to enable the focal point to move along a desired path, the position at each time instant k is designed to be $\boldsymbol{p}^*(k)$ with a corresponding $\boldsymbol{\varphi}^*(k)$ satisfying

$$\boldsymbol{p}^*(k+1) = f\left(\boldsymbol{p}^*(k), \boldsymbol{\varphi}^*(k)\right). \tag{5}$$

As focal position could be sensed by intensity maps, the error system with the state $\boldsymbol{e}(k) = \boldsymbol{p}(k) - \boldsymbol{p}^*(k)$ is given below:

$$\boldsymbol{e}(k+1) = f\left(\boldsymbol{p}(k), \boldsymbol{\varphi}(k), \boldsymbol{\omega}(k)\right) - f\left(\boldsymbol{p}^*(k), \boldsymbol{\varphi}^*(k)\right). \tag{6}$$

Based on (5) and (6), the following characterization of the state, control input, and disturbance is given

$$\begin{aligned} &f\left(\boldsymbol{p}(k), \boldsymbol{\varphi}(k), \boldsymbol{\omega}(k)\right) - f\left(\boldsymbol{p}^*(k), \boldsymbol{\varphi}^*(k)\right) \\ &= \boldsymbol{A}(k)\boldsymbol{e}(k) + \boldsymbol{B}_u(k)\left(\boldsymbol{\varphi}(k) - \boldsymbol{\varphi}^*(k)\right) + \boldsymbol{B}_\omega(k)\boldsymbol{\omega}(k). \end{aligned} \tag{7}$$

With state feedback control input $\boldsymbol{\varphi}(k) = \boldsymbol{K}(k)\boldsymbol{e}(k) + \boldsymbol{\varphi}^*(k)$, the following condition is given to design the control matrix at instant k and minimize the effect of disturbance:

$$\begin{bmatrix} \left(\boldsymbol{A}^T + \boldsymbol{K}^T \boldsymbol{B}_u^T\right)\boldsymbol{P}\left(\boldsymbol{A} + \boldsymbol{B}_u\boldsymbol{K}\right) - \boldsymbol{P} + \boldsymbol{I} & \boldsymbol{P}\left(\boldsymbol{A} + \boldsymbol{B}_u\boldsymbol{K}\right) \\ \left(\boldsymbol{P}\left(\boldsymbol{A} + \boldsymbol{B}_u\boldsymbol{K}\right)\right)^{\mathrm{T}} & \boldsymbol{B}_\omega^T \boldsymbol{B}_\omega - \gamma^2 \boldsymbol{I} \end{bmatrix} < 0, \tag{8}$$

where γ is the output tracking performance of the system with disturbance. The relationship between tracking error and disturbance can be represented as the equation

$$\|\boldsymbol{e}\|_2 < \gamma \|\boldsymbol{\omega}\|_2 . \tag{9}$$

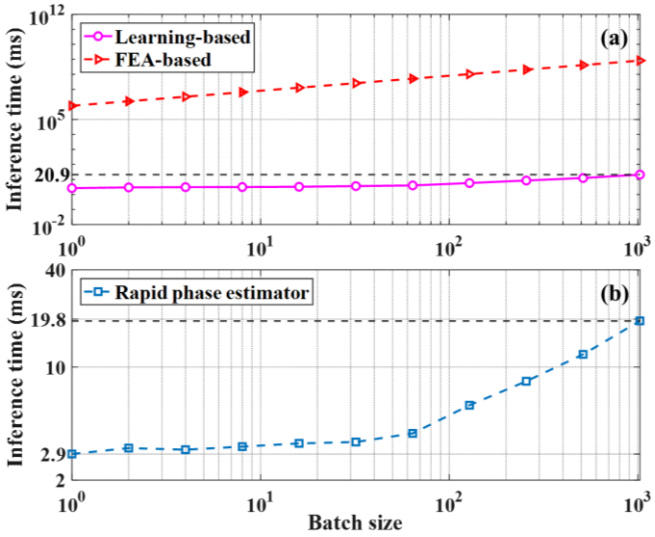

Fig. 4. Computational efficiency of **(a)** the proposed learning-based modulation model and conventional FEA-based method, as well as **(b)** the phase estimator. The processing time is shown in a logarithmic scale.

D. Implementation Details

As a proof-of-concept, we first conducted acoustic simulations to collect raw data for model training. Acoustic fields were computed using finite element analysis (FEA) in Comsol Multiphysics and Matlab co-simulator (Comsol Inc). A high-resolution axial MR image of the abdomen was acquired and segmented for simulations [29]. The liver, as the biggest solid organ in the abdomen, was demonstrated as a representative organ for experiments. Since the framework was designed for a wide range of applications, a medium frequency (i.e., 550 kHz) was chosen. Based on the time-reversal principle [17], for each sample, a point element is set at the target to transmit ultrasound waves in all directions in the time domain. The transducer array acting as receivers recorded corresponding time-of-flight (ToF) and acoustic pressure (i.e., amplitude pattern). By transforming the ToF to the phase in [0, 2π], we could obtain the dataset consisting of the target position, phase pattern, and amplitude pattern.

Then, we evaluated the performance of the model and the controller. With predicted phase and amplitude patterns, forward wave propagation simulation in the frequency domain was performed to visualize the acoustic intensity distribution. A homogenous scenario (i.e., water) without aberrations, i.e., the non-corrected (NC) case, was involved for comparisons with refocusing results of our model. Since medium heterogeneity was not considered, the geometric solution was used to calculate the phase of the i^{th} element:

$$\varphi_i = \frac{\pi}{180T} \operatorname{mod}\left(\frac{\left\|\boldsymbol{r}_i - \boldsymbol{p}^*\right\|}{c_0}, T\right), \tag{10}$$

where $\boldsymbol{r}_i$ depicts the i^{th} element's position, T is the ultrasound propagation cycle, and c_0 is the acoustic speed in water. Targeting accuracy was calculated with the displacement from the focal point and the target position. To calculate the focal obliquity, the acoustic intensity distribution was normalized, which can be characterized by a 2D Gaussian distribution [18]:

$$G = \frac{1}{2\pi}\left(\det\left(\boldsymbol{M}\right)\right)^{-0.5} \exp\left(-0.5\left(\boldsymbol{x} - \boldsymbol{\mu}\right)\boldsymbol{M}^{-1}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)^{\mathrm{T}}\right), \tag{11}$$

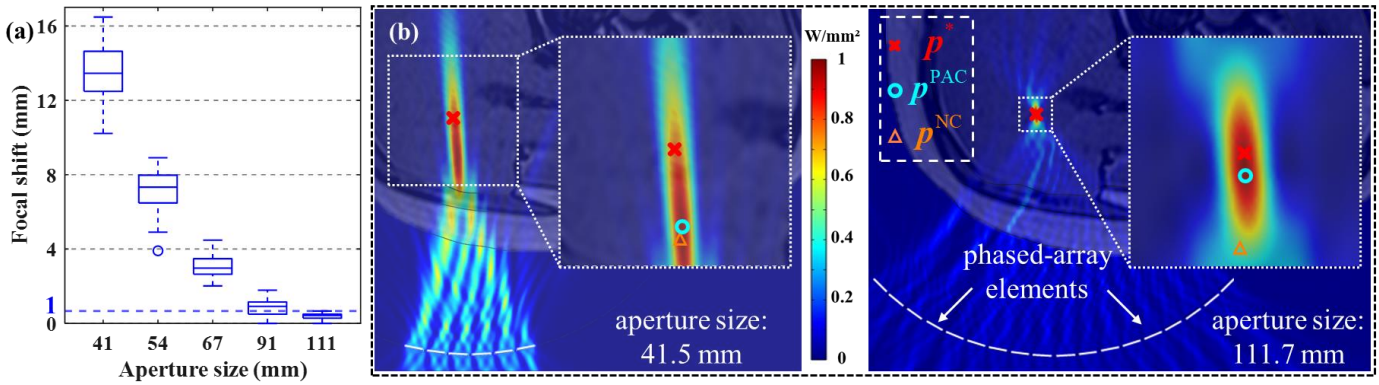

Fig. 5. (a) Transducer aperture size *w.r.t.* focal shift**. (b)** Two examples showing effect of aperture size on beam forming. The focal spots are visualized with normalized intensity within the liver ROI. $\boldsymbol{p}^{*}$, $\boldsymbol{p}^{\mathrm{PAC}}$, and $\boldsymbol{p}^{\mathrm{NC}}$ denote the focal positions under the ground truth, phase-amplitude correction, and non-corrected conditions, respectively.

where exp(·) denotes the exponential function, $\boldsymbol{\mu}$ represents the position with peak intensity, $\boldsymbol{x}$ stands for a position in the task space, and $\boldsymbol{M} \in \mathbb{R}^{2\times 2}$ is a weighted covariance matrix:

$$\boldsymbol{M} = \frac{\sum w_j}{\left(\sum w_j\right)^2 - \sum w_j^2} \sum w_j \left(\boldsymbol{x}_j - \boldsymbol{\mu}\right)^{\mathrm{T}} \left(\boldsymbol{x}_j - \boldsymbol{\mu}\right), \quad (12)$$

where w_j corresponds to the intensity value at j^{th} position $\boldsymbol{x}_j$. Thus, the focal obliquity $\boldsymbol{\theta}$ can be calculated by the angle between the transducer's central axis and the covariance matrix's eigenvector with the largest eigenvalue.

III. Experiments, Results, and Discussion

A. Model Sensitivity Analysis

Prior to model training, the number of samples within a predefined ROI has to be considered. The model sensitivity has to be analyzed to quantify the relationship between model accuracy and sample size, in order to verify whether a small dataset is still sufficient for effective modulation, thus alleviating the labor-intensive data collection process.

Here, a rectangle ROI with 40 mm wide and 20 mm long is defined. Such a wide range could be sufficient to validate our model's and controller's performances. The minimum sampling space was 1 mm, thus obtaining a total of 649 training samples fully covering the whole ROI. Thirty testing samples were selected randomly for prediction. By enlarging the sampling spacing, the sample size can be decreased. We tuned the sampling spacing to enable the sample size to decrease from 100% to 10% at an interval of 10%. Then, the modulation model was trained with the corresponding dataset. The mean error of the testing samples in terms of the phase and amplitude were calculated individually, such that the relationship between the phase/amplitude prediction accuracy and sample size can be quantified. The total sample size leads to a mean error of 1.4˚ and 0.17 Pa in phase (**Fig. 3a**) and amplitude (**Fig. 3b**) modulation, respectively, indicating that 1-mm sampling spacing is sufficient for robust prediction. A notable trade-off can be seen between accuracy and sample size. The half bandwidth of the phase error reaches only 20% sample size, corresponding to 14.1˚ within the bandwidth. Similarly, 30% of the sample size could result in a half decrease in amplitude modulation accuracy, at which the mean error is kept at 10.3 Pa. Such wide bandwidth of the two patterns demonstrates promising applicability for real applications with particular accuracy criteria, greatly reducing the computational costs with large amounts of samples.

B. Computational Efficiency

As thousands of points have to be targeted in real applications, aberrations correction would be a time-consuming step during the whole treatment process. Thus, our model's computational efficiency with respect to (*w.r.t.*) batch size was tested to evaluate its feasibility in clinical usage. Ten rounds of the computation time were averaged and recorded. The traditional FEA-based simulation method was included for analysis. The average time was recorded for 10 simulations. The simulation and model computation were performed on a server (AMD Ryzen Threadripper 3960X).

Taking full-wave propagation into account, the numerical simulation method is computationally intensive, requiring around 810,000 ms for correction on only one target, and costing 230.4 hours for 1024 targets (**Fig. 4a**). Such a long correction time may be unacceptable in a clinical standpoint. Instead, our model enables rapid (2.65 ms per target), mesh-free, and accurate (mean error: 1.39˚) full-wave aberrations compensation. Moreover, it could complete modulation for 1024 targets within a short time (20.89 ms), indicating that our learning-based compensation strategy is applicable in clinical environment. The computational efficiency of the phase estimator was also tested (**Fig. 4b).** Estimation on one target only requires 2.9 ms, maintaining a high update frequency (>100 Hz) for closed-loop control. Such a rapid estimation could provide a computationally efficient solution, allowing the control policy for real-time tracking.

C. Machine-learning-enabled Transducer Design

Design and optimization of the transducer configuration is a major concern in clinical usage. Acoustic simulations were performed for compact design prior to real applications. Several parameters were determined first. Owing to inherent focusing capability, transducer geometry was defined as a spherical shell. As its radius is associated with the maximum focal depth, it was set at 75 mm. The element size in diameter of Ø5.5 mm was determined by the frequency (550 kHz). The objective is to optimize the aperture size for a compact design with requirements of focal depth (max. 85 mm) and steering range of 10×10 mm. The beam is only steered along an "+" shaped path, i.e., sequentially targets 10 points with 1-mm spacing in both axial and lateral directions.

As a preliminary study, we employed our modulation model to guide transducer design in regard to aperture size. Given a particular aperture size, the model acts as a fast solver to ensure efficient transduction. With a larger aperture size, focal shift and area are continuously reduced (**Fig. 5a**). When the aperture is 111.7 mm, the maximum error (<1 mm) is within clinical requirements [27]. For all the targets with an intensity of -3 dB, their area differences are within 3.1 mm^2, which is negligible. These results indicate that the optimal number of elements is 18. Such a compact transducer could accommodate the required steering range and focal depth. In **Fig. 5b**, the elements on a spherical surface are extended from the transducer axis outward. Forward ultrasound propagation simulation was performed, showing the beam interference with heterogeneous tissues based on the MR image. Despite accurate phase modulation, the aperture size with 6 elements results in superficial beam penetration, causing insufficient focal depth. With a larger aperture size, the beams can be focused deeper inside the liver with a much smaller focal area. The f-number of the optimized transducer is smaller than 1.0, showing potential to minimize near-field overheating, which is consistent with previous study [30]. Our ML-enabled method enables rapid modulation to significantly reduce the time required for design. This could not only efficiently determine the appropriate transducer size, but also balance the trade-off between computational cost and fabrication cost. It is worth noting that the conventional design relies on repeated time-reversal-based simulation, such that optimal design can be a time-consuming, computational-costing, and labor-intensive process, particularly for the transducer with thousands of tightly spaced elements. In contrast, phases of unlimited elements at arbitrary positions can be collected for our model training. Once trained, the model can be repeatedly employed for design in particular applications.

D. Focal Targeting with Phase-Amplitude Correction

Provided with sufficient focal depth and workspace, the targeting accuracy of our modulation model has to be evaluated, so as to test the feasibility of steering the focal spot to "fill" the ROI. Obvious tissue heterogeneity can be seen in the MR image (**Fig. 6a**), where the sound speed of two soft tissues may vary by over 10%. The designed compact transducer is placed 20 mm beneath the skin surface. Twenty-five evenly distributed points with a 2-mm spacing constitute a rectangle ROI, starting from #1 and ending at #25.

Table 1. Mean and max. of focusing accuracy.

	NC	AC	PAC	**Improv.**
Mean±std	0.84±0.26	0.80±0.27	0.35±0.21	**58.3%**
Max.	1.48	1.48	0.65	**56.1%**

Improv. denotes improvement of the PAC condition over the NC condition.

The influence of the NC (i.e., baseline), amplitude-only correction (AC), and phase-amplitude correction (PAC) in targeting accuracy were studied (**Table 1**). The AC case contributes focal targeting marginally, which is similar to the NC case. Instead, the PAC case could "strengthen" the targeting capability by modulating both patterns simultaneously. It achieves a 56.1% improvement compared with the baseline (1.48 mm), meeting clinical requirement [27]. Besides, the PAC case results in only a 0.35 mm error with a 58.3% improvement over the baseline, showing effective modulation performance.

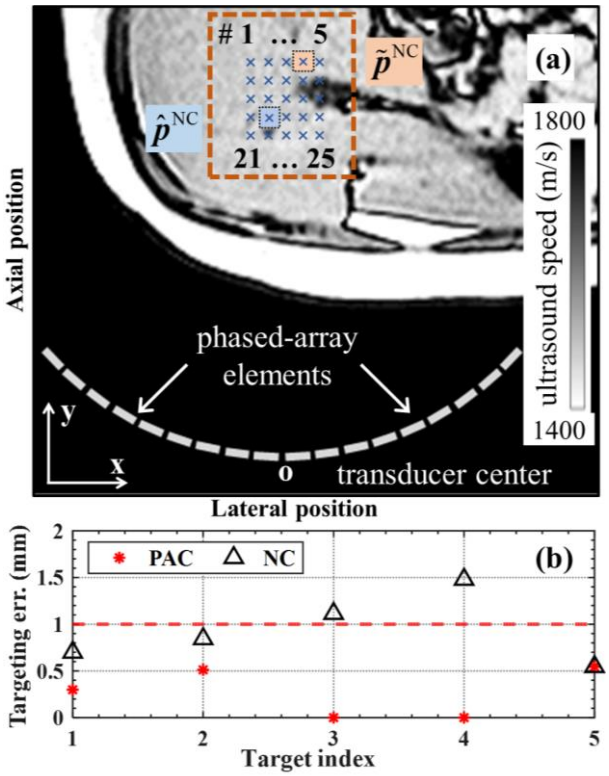

Fig. 6. (a) Testing samples (#1~25) selected in the liver ROI**.** The superscripts of $\boldsymbol{p}$, "~" and "^", respectively, denote the target with max. (#4) and min. (#17) focal targeting accuracy improvement. **(b)** Focal targeting accuracy at the five deepest target points.

Note that off-target heating remains a major concern in HIFU owing to insufficient accuracy. Such a small mean error demonstrates our model's potential to reduce off-target heating to improve treatment safety. It also indicates that it can balance the trade-off between correction accuracy and simulation time.

Our primary concern is the deep-seated targets, as beams would have to pass through many layers of tissues, suffering from severe aberrations, making effective beams convergence rather challenging. Thus, given the estimated phases, the 5 deepest targets were chosen for evaluation (**Fig. 6a**). The focal spot was refocused at the targets successively along the lateral direction. It is obvious that targeting at all the points leads to slight errors, some of which are negligible (**Fig. 6b**). This not only suggests that our method ensures robust targeting of the deep tissues, but also indicates that the optimized transducer is applicable for HIFU targeting.

E. Focal Obliquity Assessment

Apart from focal shift, focal obliquity is another crucial indicator in refocusing quality. To ensure safe HIFU ablation, the prerequisites are that obliquity should be minimized, and grating or side lobes need to be suppressed. Note that those lobes are unwanted and would cause damage to normal tissues. Here, the objective is to investigate the effect of our modulation model in focal obliquity alignment and lobes reduction. Four conditions are considered, i.e., NC, AC, phase-only correction (PC), and PAC cases. NC case indicates that equal deposition of the element's power is applied.

The focal obliquity of the 25 testing targets (**Fig. 6a)** is shown in **Table 2**. NC results in the largest obliquity (mean: 8.4˚). Although AC contributes to limited focal alignment (mean: 8.0˚), its conjunction with PC significantly reduces

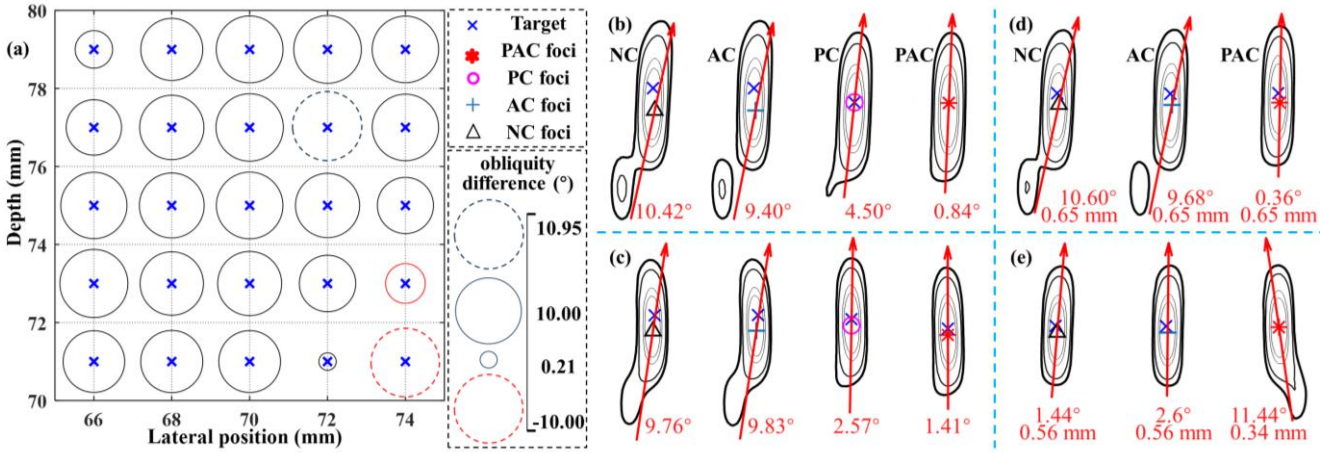

Fig. 7. (a) Focal obliquity difference between PAC and NC cases, where the max. and min. differences are represented by bubbles with black and red dashed lines, respectively. Beam profiles of the targets with **(b)** max. (#4) and **(c)** min. (#17) targeting accuracy, and the targets with **(d)** max. (#9) and **(e)** min. (#25) obliquity reduction. The red arrow represents the covariance matrix's eigenvector with the largest eigenvalue.

obliquity to 2.7˚. It is worth noting that such a combination accomplishes a 67.9% improvement over NC, demonstrating that our model may substantially improve HIFU safety. It is clinically significant, particularly for ROI surrounded by sensitive structures (e.g., nerves, blood vessels).

Table 2. Mean and max. of focal obliquity.

Obliquity (deg)	NC	AC	PC	PAC	Improv.
Mean	8.4	8.0	3.2	2.7	**67.9%**
Max.	11.5	12.0	9.1	11.4	/

Improv. denotes improvement of the PAC condition over the NC condition.

The focal obliquity difference between PAC and NC is shown in **Fig. 7a.** Although increasing the shifted distance would cause an unpredictable increase in obliquity [31], making alignment challenging for those targets far from the natural foci, our model still reduces the obliquity of most focal spots significantly, even for all the five deepest targets. With PAC correction, target #4 (**Fig. 7b**) and target #17 (**Fig. 7c**) attain max. (0 mm) and min. (0.32 mm) targeting accuracy improvement, respectively. In the NC case, the side lobe is beside the main lobe, resulting in unnecessary tissue injury. The AC case is unstable in obliquity alignment, where the grating lobe may appear (#4), causing unexpected outer damage. However, its combination with PC (i.e., PAC) would greatly alleviate obliquity, ensuring negligible focal orientation in both targets. The consistent finding in these two targets not only suggests our model's effectiveness in focal alignment, but also indicates that focal targeting is not significantly correlated with focal obliquity. Focal obliquity can be another indicator of beam refocusing quality.

The targets with max. (#9, **Fig. 7d**) and min. (#25, **Fig. 7e**) obliquity differences between NC and PAC are also analyzed. Similar to #17, despite acceptable targeting accuracy (0.65 mm) under NC condition, the obliquity is still relatively large (10.6˚) with obvious side lobes. The possible reason is that the sidelobes would induce the obliquity. Regarding PAC, the distorted wavefront is corrected, suggesting optimization on elements' power deposition can substantially "strengthen" the main lobe (i.e., focal spot) and obliterate grating/side lobes. For target #25, this shallow target has a small focal shift even without correction, but either AC or PAC deteriorates focal alignment, even resulting in a side lobe in the PAC case. This is possible that the large incident angles (>25˚) of some elements may cause strong reflection and lead to ineffective energy deposition at the target [32]. Robotic manipulation could avoid this problem by adjusting the transducer orientation. It is worth noting that no grating lobes appear in all four targets under the NC case, showing robust beam directivity within the ROI. This implies that our designed array is effective in beam refocusing.

F. Continuous Path Tracking

This experiment aims to study the effectiveness of robust H_∞ control for path tracking in the presence of unknown disturbances. In this regard, the phase estimator was initially examined. Given the nonlinear nature of the function, it is necessary to determine matrices $\boldsymbol{A}$, $\boldsymbol{B}_u$, $\boldsymbol{B}_\omega$ in (7) to characterize the nonlinear function. Subsequently, a system is established, with position $\boldsymbol{p}(k)$ representing the state and phase $\boldsymbol{\varphi}(k)$ acting as the control input. By introducing disturbances into the constructed simulation platform (3), the actual output is simulated under open-loop conditions. Here, it is assumed that the system disturbance is characterized by white noise, following a normal distribution $N(0,0.2)$ in the interval of the time step [1,16], in order to emulate the subtle, irregular, and slight movements of the human body. The focal point adheres to a pre-established rectangular trajectory $\boldsymbol{p}^*(k)$ measuring 12×20 mm, advancing at a step size of 1 mm throughout the trajectory. As a result of this disturbance, we have found that the average error for the open loop system is 0.895 mm. Additionally, we have identified the point at which the error is at its maximum, which occurs at the 17th step. The robust controller is implemented by utilizing information from the phase and focal point model and the detected focal point position. The feedback matrix $\boldsymbol{K}$ is determined by minimizing γ in (8). Based on the designed time-varying control matrix function, it can be observed that the error is reduced compared to the open-loop system, with approximately 0.3 times smaller values. The findings reveal that our robust H_∞ controller is capable of managing unknown disturbances, and maintaining a path tracking error in an acceptable range.

IV. Conclusions and Future Work

In this paper, we present a learning-based modulation and control framework for widespread MRg-FUS treatments

without prior knowledge of heterogeneous media. The model ensures rapid (2.65 ms) full-wave correction on phase and amplitude aberrations simultaneously. This enables high targeting accuracy (mean 0.35 mm, max. 0.65 mm) under the clinical requirement [27], thus alleviating off-target heating. Besides, the unwanted grating lobes can be effectively suppressed, and the focal obliquity is "aligned" to only 2.7°, thus reducing the risk of tissue damage. Compared with the NC case, 58.3% targeting error and 67.9% obliquity is decreased. Accrediting to rapid modulation, our ML-enabled approach significantly reduces the time required for optimal transducer design, compared with the labor-intensive and computational-costing conventional method. Results from model sensitivity and computational efficiency also support the model's feasibility in HIFU over abdominal organs.

The closed-loop hybrid control approach ensures accurate tracking accuracy (0.30 mm) along an intended trajectory, even under unknown disturbance. Such robust tracking performance demonstrates our control policy's adaptability inside the human body in LIFU applications. The estimator can offer a computationally efficient phase actuation (<3 ms) for the robust controller to facilitate real-time beam motion control. In the future, we will implement proof-of-concept transduction and controller in a series of *ex vivo* or *in vivo* HIFU/LIFU experiments (e.g., mild hyperthermia). Besides, amplitude correction may cause increased power emissions to those healthy tissues with high absorption, thus resulting in the risk of overheating. Utilizing the ML-powered design method to further optimize the transducer configuration (e.g., saddle-shaped transducer) will be a promising solution.

REFERENCES

[1] M. A. Santos, S.-K. Wu, M. Regenold, C. Allen, D. E. Goertz, and K. Hynynen, "Novel fractionated ultrashort thermal exposures with MRI-guided focused ultrasound for treating tumors with thermosensitive drugs," *Science Advances,* vol. 6, no. 36, p. eaba5684, 2020.

[2] J. Dai *et al.*, "A robotic platform to navigate MRI-guided focused ultrasound system," *IEEE Robotics and Automation Letters,* vol. 6, no. 3, pp. 5137-5144, 2021.

[3] Y. Meng, K. Hynynen, and N. Lipsman, "Applications of focused ultrasound in the brain: from thermoablation to drug delivery," *Nature Reviews Neurology,* vol. 17, no. 1, pp. 7-22, 2021.

[4] Q. Wang *et al.*, "Ultrasound Doppler-guided real-time navigation of a magnetic microswarm for active endovascular delivery," *Science Advances,* vol. 7, no. 9, p. eabe5914, 2021.

[5] Q. Wang and L. Zhang, "External power-driven microrobotic swarm: from fundamental understanding to imaging-guided delivery," *ACS Nano,* vol. 15, no. 1, pp. 149-174, 2021.

[6] A. Gunderman, R. Montayre, A. Ranjan, and Y. Chen, "Review of Robot-Assisted HIFU Therapy," *Sensors,* vol. 23, no. 7, p. 3707, 2023.

[7] G. Fang *et al.*, "Soft robotic manipulator for intraoperative MRI-guided transoral laser microsurgery," *Science Robotics,* vol. 6, no. 57, p. eabg5575, 2021.

[8] Z. Dong *et al.*, "High-performance continuous hydraulic motor for MR safe robotic teleoperation," *IEEE Robotics and Automation Letters,* vol. 4, no. 2, pp. 1964-1971, 2019.

[9] K. Kim, E. Breton, A. Gangi, and J. Vappou, "Simultaneous fat-referenced proton resonance frequency shift thermometry and MR elastography for the monitoring of thermal ablations," *Magnetic Resonance in Medicine,* vol. 84, no. 1, pp. 339-347, 2020.

[10] J. Vappou, P. Bour, F. Marquet, V. Ozenne, and B. Quesson, "MR-ARFI-based method for the quantitative measurement of tissue elasticity: application for monitoring HIFU therapy," *Physics in Medicine & Biology,* vol. 63, no. 9, p. 095018, 2018.

[11] K. Choquet, J. Vappou, P. Cabras, O. Ishak, A. Gangi, and E. Breton, "Magnetic Resonance Acoustic Radiation Force Imaging (MR-ARFI) for the monitoring of High Intensity Focused Ultrasound (HIFU) ablation in anisotropic tissue," *Magnetic Resonance Materials in Physics, Biology and Medicine,* pp. 1-11, 2023.

[12] Y. Qiao *et al.*, "Simultaneous acoustic radiation force imaging and MR thermometry based on a coherent echo-shifted sequence," *Quantitative Imaging in Medicine and Surgery,* vol. 10, no. 9, p. 1823, 2020.

[13] V. Auboiroux *et al.*, "ARFI-prepared MRgHIFU in liver: simultaneous mapping of ARFI-displacement and temperature elevation, using a fast GRE-EPI sequence," *Magnetic Resonance in Medicine,* vol. 68, no. 3, pp. 932-946, 2012.

[14] C. Yiallouras, K. Ioannides, T. Dadakova, M. Pavlina, M. Bock, and C. Damianou, "Three-axis MR-conditional robot for high-intensity focused ultrasound for treating prostate diseases transrectally," *Journal of Therapeutic Ultrasound,* vol. 3, pp. 1-10, 2015.

[15] E. Epaminonda, T. Drakos, C. Kalogirou, M. Theodoulou, C. Yiallouras, and C. Damianou, "MRI guided focused ultrasound robotic system for the treatment of gynaecological tumors," *The International Journal of Medical Robotics and Computer Assisted Surgery,* vol. 12, no. 1, pp. 46-52, 2016.

[16] C. R. Dillon, A. Farrer, H. McLean, S. Almquist, D. Christensen, and A. Payne, "Experimental assessment of phase aberration correction for breast MRgFUS therapy," *International Journal of Hyperthermia,* vol. 34, no. 6, pp. 731-743, 2018.

[17] A. Kyriakou, E. Neufeld, B. Werner, M. M. Paulides, G. Szekely, and N. Kuster, "A review of numerical and experimental compensation techniques for skull-induced phase aberrations in transcranial focused ultrasound," *International Journal of Hyperthermia,* vol. 30, no. 1, pp. 36-46, 2014.

[18] A. Hughes, Y. Huang, A. Pulkkinen, M. L. Schwartz, A. M. Lozano, and K. Hynynen, "A numerical study on the oblique focus in MR-guided transcranial focused ultrasound," *Physics in Medicine & Biology,* vol. 61, no. 22, p. 8025, 2016.

[19] L. Marsac *et al.*, "Ex vivo optimisation of a heterogeneous speed of sound model of the human skull for non-invasive transcranial focused ultrasound at 1 MHz," *International Journal of Hyperthermia,* vol. 33, no. 6, pp. 635-645, 2017.

[20] C. Jin, D. Moore, J. Snell, and D.-G. Paeng, "An open-source phase correction toolkit for transcranial focused ultrasound," *BMC Biomedical Engineering,* vol. 2, no. 1, pp. 1-11, 2020.

[21] S. Schoen and C. D. Arvanitis, "Heterogeneous angular spectrum method for trans-skull imaging and focusing," *IEEE Transactions on Medical Imaging,* vol. 39, no. 5, pp. 1605-1614, 2019.

[22] M. Chen *et al.*, "Numerical and experimental evaluation of low-intensity transcranial focused ultrasound wave propagation using human skulls for brain neuromodulation," *Medical Physics,* vol. 50, no. 1, pp. 38-49, 2023.

[23] C. Mougenot, M. Tillander, J. Koskela, M. O. Köhler, C. Moonen, and M. Ries, "High intensity focused ultrasound with large aperture transducers: a MRI based focal point correction for tissue heterogeneity," *Medical Physics,* vol. 39, no. 4, pp. 1936-1945, 2012.

[24] X. Wang *et al.*, "Learning-Based Visual-Strain Fusion for Eye-in-Hand Continuum Robot Pose Estimation and Control," *IEEE Transactions on Robotics,* 2023.

[25] X. Wang, Y. Li, and K.-W. Kwok, "A survey for machine learning-based control of continuum robots," *Frontiers in Robotics and AI,* vol. 8, p. 730330, 2021.

[26] K. Wang *et al.*, "Large-Scale Surface Shape Sensing with Learning-Based Computational Mechanics," *Advanced Intelligent Systems,* vol. 3, no. 11, p. 2100089, 2021.

[27] S. Schoen Jr *et al.*, "Towards controlled drug delivery in brain tumors with microbubble-enhanced focused ultrasound," *Advanced Drug Delivery Reviews,* vol. 180, p. 114043, 2022.

[28] G. Huang, G.-B. Huang, S. Song, and K. You, "Trends in extreme learning machines: A review," *Neural Networks,* vol. 61, pp. 32-48, 2015.

[29] A. E. Kavur *et al.*, "CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation," *Medical Image Analysis,* vol. 69, p. 101950, 2021.

[30] V. A. Khokhlova *et al.*, "Design of HIFU transducers to generate specific nonlinear ultrasound fields," *Physics Procedia,* vol. 87, pp. 132-138, 2016.

[31] H. Q. Lean and Y. Zhou, "Acoustic field of phased-array ultrasound transducer with the focus/foci shifting," *Journal of Medical and Biological Engineering,* vol. 39, pp. 919-931, 2019.

[32] H. Baek *et al.*, "Clinical intervention using focused ultrasound (FUS) stimulation of the brain in diverse neurological disorders," *Frontiers in Neurology,* vol. 13, p. 880814, 2022.