Real-Time Passive Acoustic Mapping Using Sparse Matrix Multiplication

Hermes A. S. Kamimura, *Member, IEEE*, Shih-Ying Wu, *Member, IEEE*, Julien Grondin, *Member, IEEE*, Robin Ji, *Graduate Student Member, IEEE*, Christian Aurup, *Graduate Student Member, IEEE*, Wenlan Zheng, Marc Heidmann, Antonios N. Pouliopoulos, *Member, IEEE*, and Elisa E. Konofagou, *Senior Member, IEEE*

***Abstract*— Passive acoustic mapping enables the spatiotemporal monitoring of cavitation with circulating microbubbles during focused ultrasound (FUS)-mediated blood–brain barrier opening. However, the computational load for processing large data sets of cavitation maps or more complex algorithms limit the visualization in real-time for treatment monitoring and adjustment. In this study, we implemented a graphical processing unit (GPU)-accelerated sparse matrix-based beamforming and time exposure acoustics in a neuronavigation-guided ultrasound system for real-time spatiotemporal monitoring of cavitation. The system performance was tested *in silico* through benchmarking, *in vitro* using nonhuman primate (NHP) and human skull specimens, and demonstrated *in vivo* in NHPs. We demonstrated the stability of the cavitation map for integration times longer than 62.5 μs. A compromise between real-time displaying and cavitation map quality obtained from beamformed RF data sets with a size of 2000 × 128 × 30 (axial pixels × lateral pixels × samples) was achieved for an integration time of 1.44 μs, which required a computational time of 0.27 s (frame rate of 3.7 Hz) and could be displayed in real-time between pulses at PRF = 2 Hz. Our benchmarking tests show that the GPU sparse-matrix algorithm processed the RF data set at a computational rate of 0.03 ± 0.01 μs/pixel/sample, which enables adjusting the frame rate and the integration time as needed. The neuronavigation system with real-time implementation of cavitation mapping facilitated the localization of the cavitation activity and helped to identify distortions due to FUS phase aberration. The *in vivo* test of the method demonstrated the feasibility of GPU-accelerated sparse matrix computing in a close to a clinical condition, where focus distortions exemplify problems during treatment. These experimental conditions show the need for spatiotemporal monitoring of cavitation with real-time capability that enables the operator to correct or halt the sonication in case substantial aberrations are observed.**

Manuscript received May 11, 2020; accepted June 7, 2020. This work was supported by the National Institutes of Health under Grant R01AG038961, Grant R01EB009041

(Hermes A. S. Kamimura and Shih-Ying Wu contributed equally to this work.) (Corresponding author: Elisa E. Konofagou.)

Hermes A. S. Kamimura, Julien Grondin, Robin Ji, Christian Aurup, Wenlan Zheng, Marc Heidmann, Antonios N. Pouliopoulos, and Elisa E. Konofagou are with the Department of Biomedical Engineering, Columbia University, New York, NY 10027 USA (e-mail: ek2191@columbia.edu).

Shih-Ying Wu was with the Department of Biomedical Engineering, Columbia University, New York, NY 10027 USA. She is now with the Information Services Department, Boston Children's Hospital, Boston, MA 02115 USA.

Digital Object Identifier 10.1109/TUFFC.2020.3001848

***Index Terms*— Drug delivery, graphical processing unit (GPU)-acceleration, nonhuman primate (NHP), passive acoustic mapping (PAM), sparse matrix, ultrasound-mediated blood–brain barrier (BBB) opening.**

I. Introduction

FOCUSED ultrasound (FUS) can drive microbubble-seeded cavitation that enhances drug delivery through the blood–brain barrier (BBB)—a semipermeable structure of the brain vasculature that prevents drug uptake into the central nervous system [1]. FUS-induced cavitation can transiently and locally disrupt the BBB [2] via transcytosis, tight junction opening, and inhibition of active transport proteins in the brain endothelial cells [3]–[5]. Preclinical studies have demonstrated the potential of FUS-mediated BBB opening to deliver variable-sized molecules such as antibody-based anticancer agents [6]–[8], antiamyloid antibodies [6], [9], [10], brain-derived neurotrophic factor [11]–[13], adeno-associated viruses [14], [15], and stem cells [16]. Currently, clinical studies are assessing the safety and feasibility of the technique for the treatment of Alzheimer's disease [17] and glioblastoma [18].

Passive cavitation detection (PCD) using single-element transducers has been used to monitor potential harmful cavitation regime in real-time in open-loop and closed-loop systems [19]–[22] inside and outside the magnetic resonance imaging (MRI) scanner [23], [24]. However, single-element PCD limits the monitoring to a temporal analysis, where cavitation activity cannot be resolved spatially. Neuronavigation-guided ultrasound with real-time passive acoustic mapping (PAM) [25] can provide a high precision therapy at lower cost in comparison to magnetic resonance-guided FUS (MRgFUS) systems. The FUS neuronavigation system allows for planning trajectories toward specific brain targets that avoid pre-existent lesions, large vessels, ventricles, and other brain structures to be circumvented while PAM enables spatiotemporal monitoring of cavitation associated with BBB-opening. The spatial mapping of acoustic cavitation recorded by a multielement transducer is reconstructed using delay-and-sum (DAS) beamforming either in the time or frequency domain [26]–[36]. Altogether, this system can help detect beam aberration due to the skull [32], [37], which could be compensated by repositioning the transducer for an efficient

and safe sonication at the prescribed location, especially when multiple sonications are required for covering a larger brain volume.

Modern graphical processing units (GPUs) offer parallel processing designed for high-peak computational throughput in a short period. Previous studies have demonstrated implementations of GPU acceleration, which generated real-time visualization of microbubble activity in the brain at variable frame rates, sample integration times, number of channels, and field-of-view (FOV) (respectively, Collin *et al.* [38]: 5 Hz, 4000 samples, 32 channels, and 40 × 10 mm^2; Jones *et al.* [37]: 3.3 Hz, 200 samples, 128 channels, and $20\times20\times20$ mm^3; Lyka *et al.* [34]: 3 Hz, N/A samples, 128 channels, and 20 × 10 mm^2; Lyka *et al.* [39]: 0.25 Hz and 0. 625 mHz, 2000 samples, 128 channels, and 660 voxels; Jones *et al.* [40]: 1 Hz, 8000 samples, 256 channels, and 10 × 10 × 10 mm^3).

Programmable multiprocessing-unit architectures provide library routines optimized for sparse matrices computations, which improves real-time performance of numerical operations [41], [42]. Data processing with a sparse matrix provides higher performance than a fully sampled matrix because it eliminates operations with zero-valued elements of the matrix. In addition, sparse representation reduces data storage as it stores only the nonzero elements and their row indices. Sparse matrix operations can be accelerated even more when performed in parallel using GPU computing. Sparsity methods have been employed widely in medical imaging [43] to speed up and improve image processing and machine-learning techniques in a variety of imaging methods, such as MRI [44]–[46], digital pathology images [47], [48], computed tomography (CT) [49], [50], and ultrasound [25], [51]–[54].

In this study, we present an implementation of PAM using GPU-accelerated sparse matrix-based beamforming and time exposure acoustics (TEAs) [55], [27], which can be performed in real-time with large maps and short-duration data sets or vice versa. This study builds on prior reports by our group as indicated by Wu *et al.* [25], which employed real-time PAM focusing on the neuronavigation system implementation, where the GPU sparse-matrix implementation was not detailed or compared with other standard methods, and Hou *et al.* [51], where active mapping (harmonic motion imaging using tissue displacement tracking) was employed, but not in passive detection or for microbubble-based therapy monitoring. The novel contributions of this article are the detailed description of the implementation of the sparse matrix-based algorithm for PAM and a benchmarking comparison of the non-sparse and sparse implementations in both CPU and GPU. Tests were performed *in silico* through benchmarking, *in vitro* using skull specimens of human and nonhuman primate (NHP), as well as *in vivo* BBB opening experiments in NHP.

II. Material and Methods

A. Passive Beamforming Algorithm

The passive beamforming algorithm based on TEA [27] was implemented with a conventional 128-channel linear array imaging probe (L7-4, Philips, Bothell, WA, USA; center frequency: 5.208 MHz). A programmable ultrasound scanner (Vantage 256, Verasonics, Kirkland, WA, USA) recorded the acoustic emissions, with time t, from cavitating bubbles during sonications. The radio frequency (RF) channel data were used to reconstruct the passive cavitation maps using dynamic receive beamforming $[\sum_{n=1}^{N} S_n(r_n, r, t)]$ and then time-integrated $(\int_0^{T_i} |\cdot|^2 dt)$ over a period T_i defined as the integration time

$$C(r) = \int_0^{T_i} \left| \sum_{n=1}^{N} S_n(r_n, r, t) \right|^2 dt \quad (1)$$

$$S_n(r_n, r, t) = c_n(t + d(r_n, r)/c) \quad (2)$$

where N is the number of elements in the array, S_n is the channel data for the nth element, r_n is the location of the nth transducer element, r is the location of the pixel to be reconstructed, $c_n[t + d(r_n, r)/c]$ the received cavitation signal for the nth channel after adjusting for the time delay based on the distance between r_n and r, and c is the speed of sound. Skull-specific aberration corrections were not performed.

B. Sparse-Matrix Construction

Although GPU-based sparse matrix beamforming has been described elsewhere [25], [51], its specific implementation and application for TEA-PAM is detailed here. In our previous study, the sparse matrix algorithm was implemented using sequences with short imaging pulses. In contrast, therapeutic pulses used in BBB opening are typically composed of thousands of acoustic cycles. Therefore, the duration of received RF signals during PAM or the total amount of beamformed data is significantly larger than the active imaging described before. The DAS beamforming was accelerated using the fast sparse matrix operation performed on a GPU [Tesla K40 (real-time) or Quadro P6000 (offline processing), NVIDIA, Santa Clara, CA, USA]. All sparse matrix operations were performed in MATLAB (2017b, MathWorks, Natick, MA, USA), which has built-in GPU support for sparse matrices since version 2015a. The computing acceleration was accomplished by implementing the term $\sum_{n=1}^{N} S_n(r_n, r, t)$ from (1) using a sparse matrix multiplication followed by temporal summation of the squared beamformed RF data. The sparse matrix multiplication can be written as $y = Ax$, where the reconstructed RF data y are the result from the multiplication of A, a sparse matrix associated with the DAS operation, and x, a matrix containing the channel data reshaped into column vectors for each sample of the integration time. The number of rows in A is equal to the total number of pixels in the reconstructed image, which is the product of the lateral map size N_x and the axial map size N_z. The number of columns in A is equal to the product of the number of samples acquired per channel Z, and the number of array elements N. Table I presents all values for the parameters used in this study.

The sparse matrix was built offline prior to the experiments, as it requires minutes-to-hours of calculation. The computational time depends on the number of samples in the acquired channel data, the number of array elements N, as well as the number of pixels in the reconstructed image ($N_x \times N_z$).

TABLE I
PARAMETERS USED FOR GPU-BASED SPARSE MATRIX MULTIPLICATION IN PASSIVE ACOUSTIC MAPPING

Symbol	Parameter	Value / Units
N_x	Lateral map size	128 pixels
N_z	Axial map size	2,000 pixels
N_T	Integrated samples	10 – 2,000 samples
T_i	Integration time	1.44 – 96.1 μs
Z	Acquired RF data samples per channel	2,000 samples
N	Number of array elements	128
f_s	Sampling frequency	20.8 MHz
n	Assigned number of transducer element or channel	1-128
l	Iteration number of the standard basis vector	1-256,000

The sparse matrix was designed here using double-precision to compute images of approximately 74 mm (depth, $N_z = 2000$ axial pixels) by 38 mm (width, $N_x = 128$ lateral pixels), with Z equal to 2000 samples acquired by $N = 128$ transducer elements (or channels) at sampling frequency 20.8 MHz (4 times the receiving center frequency of the L7-4 array, which is 5.208 MHz) that can be displayed in real-time once the sparse matrix is built. The sparse matrix needs to be built offline only once and then can just be loaded into memory for real-time beamforming.

The following describes how to build the reconstruction sparse matrix numerically. First, a conventional 3-D matrix representation [Fig. 1(a)] containing the distance D in sample units between each image pixel and transducer element is calculated by

$$D = |\vec{r_n} - \vec{r}|N_w \tag{3}$$

where $\vec{r}$ is the pixel location in the reconstructed image in wavelength units, $\vec{r_n}$ is the location of the nth element in wavelength units, and N_w is the number of samples per wavelength ($N_w = 4$, yielding a sampling frequency four times the center frequency of the receiving array). The speed of sound associated with the wavelength was 1540 m/s, which is equal to the speed of sound in water.

Then, the values in the distance matrix are converted into indices associated with the size of a given data point of channel data provided by Verasonics Vantage (sample segments acquired by each transducer element per frame). The indexed distance [Fig. 1(b)] is given by

$$i(z, x, n) = D(z, x, n) + (n - 1)Z \tag{4}$$

where i is the indexed pixel to element distance in the channel data, and Z is the total depth in sample units, with $Z = 2000$ samples and $n = 1$–128. After that, the DAS operation is performed to compute the matrix $T^l(z, x)$

$$T^l(z, x) = \sum_{n=1}^{N}((1 - d(z, x, n)) \cdot B^l(i(z, x, n)) + d(z, x, n) \cdot B^l(i(z, x, n) + 1) \tag{5}$$

where l is iterated from 1 to 256 000 ($Z \times N$) and

$$d(z, x, n) = \text{mod}[D(z, x, n), 1] \tag{6}$$

where mod is the modulus operator, and B^l is the lth standard basis vector of the 256 000-dimensional Euclidian space, containing zeros everywhere except at the lth position [$B^l(i) = \delta_{il}$, where δ_{il} is the Kronecker delta, Fig. 1(c) and (d)]. Finally, the sparse matrix A is allocated with nonzero values obtained from the matrix T^l, which can be obtained in MATLAB using $[k^l, \sim, s^l] = \text{find}(T^l(:))$, where k^l is the vector of indices of nonzero values in T^l, and s^l is the corresponding vector of nonzero values in T^l. The sparse matrix [Fig. 1(e)] is given by

$$A(k^l, l) = s^l. \tag{7}$$

C. Time Exposure Acoustics Real-Time Processing

Once the sparse matrix is loaded in the computer or GPU memory, the RF data can be beamformed in real-time by simply multiplying the sparse matrix by the channel data. This is the only step in the processing where the sparse matrix is used. The acquired data sets comprise 2000 samples per receiving element, that is, a signal duration of 96.1 μs and a sampling frequency of 20.8 MHz. To acquire data sets with reduced integration times that would accelerate computation, the beamformed time-domain signal was truncated at the relevant sample after $t = 0$. For example, only the first 30 samples were used for an integration time of 1.44 μs, the first 300 samples for an integration time of 14.4 μs. An important property of the multiplication using the sparse matrix is that it can be applied regardless of the number of samples, as it refers to a multiplication involving two 2-D matrices [Fig. 2(a)]. Then, the cavitation map is obtained from the TEAs processing [(1)] using the beamformed data [Fig. 2(b)]. Finally, the image is reshaped in 2-D ($N_z \times N_x$) and displayed in real-time [Fig. 2(c)].

D. In Vitro Experiments

The *in vitro* test of the system with and without the skull specimens (NHP and human and parietal bone) was performed in a silicon phantom with a 4-mm-diameter tube where in-house, lipid-shell, monodisperse microbubbles (median diameter: 4–5 μm, diluted to 2×10^5 bubbles/mL [56], [57]) circulated at a flow rate of 0.25 mL/s using a syringe pump [Fig. 3(a)]. The skull specimens were degassed 24 h before the experiment. A customized MATLAB code controlled a single-element, 0.5-MHz FUS transducer (diameter: 64 mm and focal depth: 62.6 mm; H-107, Sonic Concepts, Bothell, WA, USA) driven by a function generator (model 33220A, Agilent Technologies, Santa Clara, CA) with 50-dB amplification (A075, ENI, NY, USA). A PCD array (L7-4, Philips, Bothell, WA, USA; center frequency: 5.208 MHz, sampling frequency: 20.8 MHz, and channel data length: 2000 samples) and a single PCD transducer (Y-107, Sonic Concepts; sensitivity: 10 kHz–15 MHz, sampling frequency: 50 MHz) were simultaneously used to monitor the cavitation generated using derated peak-negative pressure (PNP): 100–600 kPa, pulse length: 5000 cycles (10 ms), pulse repetition frequency (PRF): 10 Hz, duration: 2 s. The skull specimen was placed between the phantom and the PCD array immediately after

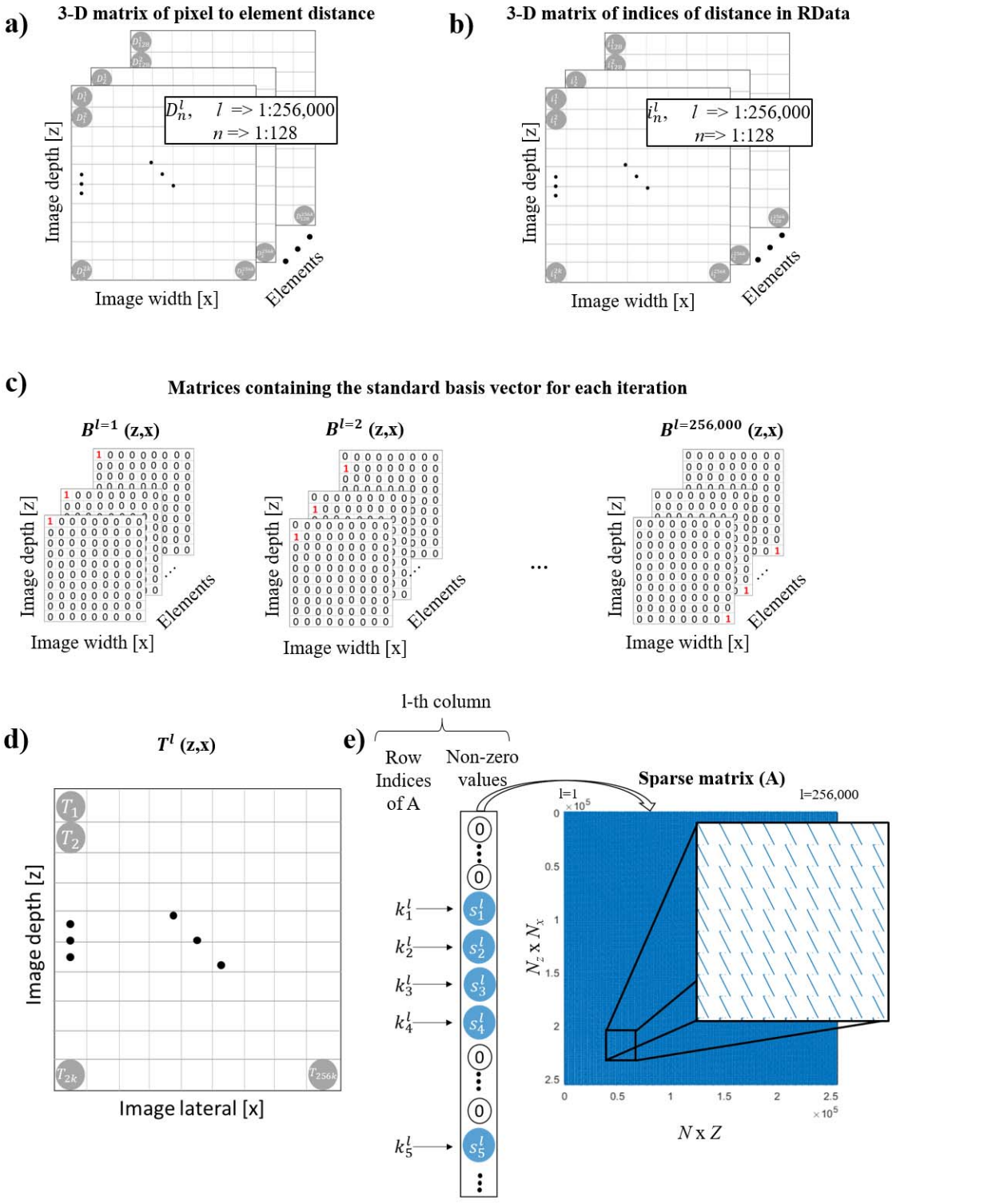


Fig. 1. Sparse matrix construction. (a) 3-D matrix of distance from the pixel to transducer element in sample units calculated for an imaging array with 128 elements, and a reconstructed image of 2000 × 128 = 256 000 pixels. (b) 3-D matrix of reindexed distance to follow data output by Verasonics Vantage. (c) Standard basis vectors used for the DAS iterative calculation. (d) Matrix resulting from DAS operation on a given standard basis vector. (e) Sparse matrix values allocation following 2-D representation.

acquisitions without the skull to assess the skull effects on the PCD data in similar experimental conditions.

E. In Vivo NHP Experiments

All procedures and experiments with animals were reviewed and approved by the Institutional Animal Care and Use Committee at Columbia University and the New York State Psychiatric Institute following the National Institutes of Health Guidelines for animal research. The *in vivo* experiments were performed in two male adult macaques (*Macaca mulatta*, weight: 9–11 kg, age: 18–20 years old). The FUS transducer was placed on the top of the animal's head using a

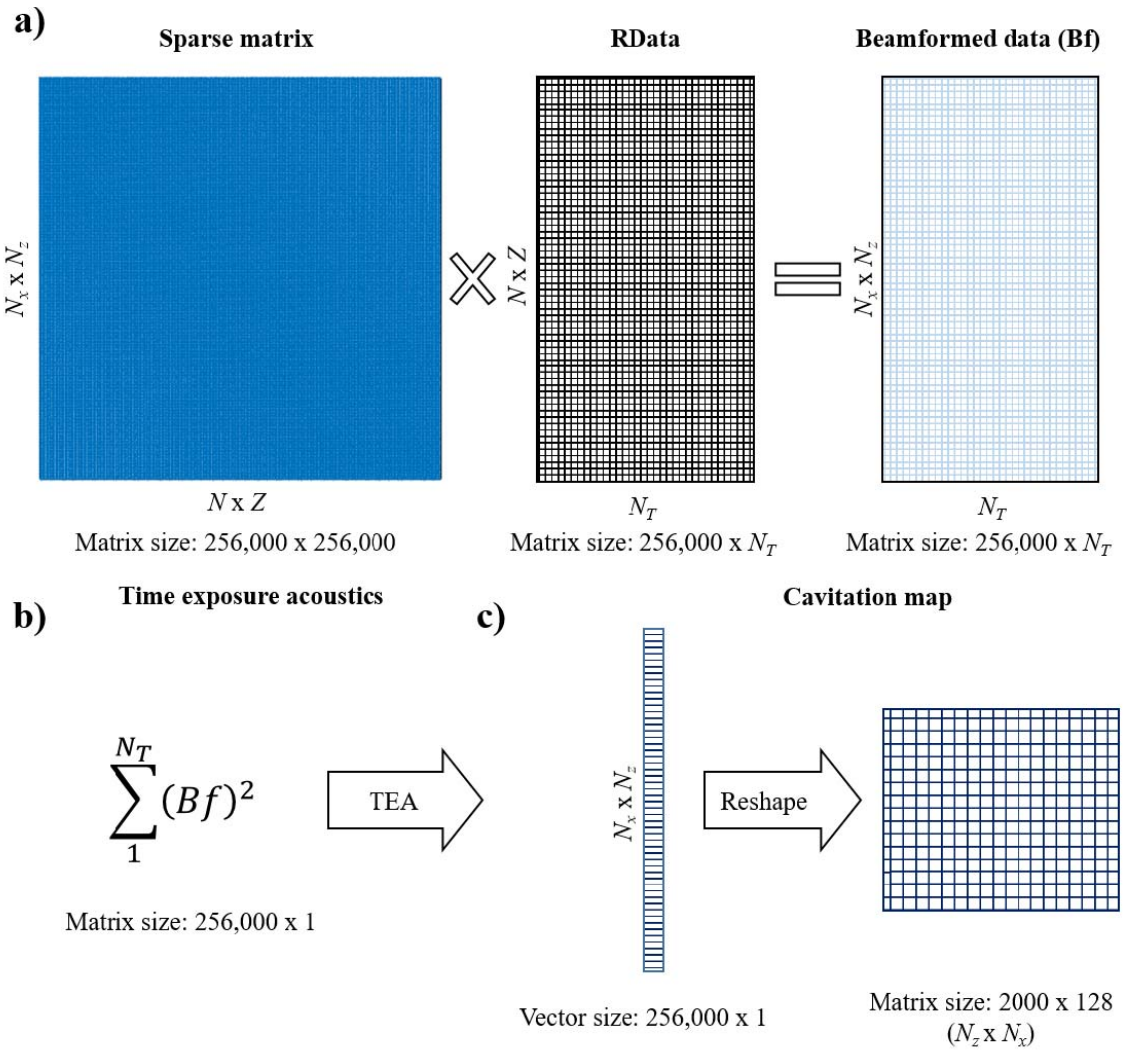


Fig. 2. TEA-PAM real-time algorithm using sparse matrix operation. (a) DAS beamforming algorithm using GPU-accelerated sparse matrix operation. (b) TEA operation. (c) Cavitation maps.

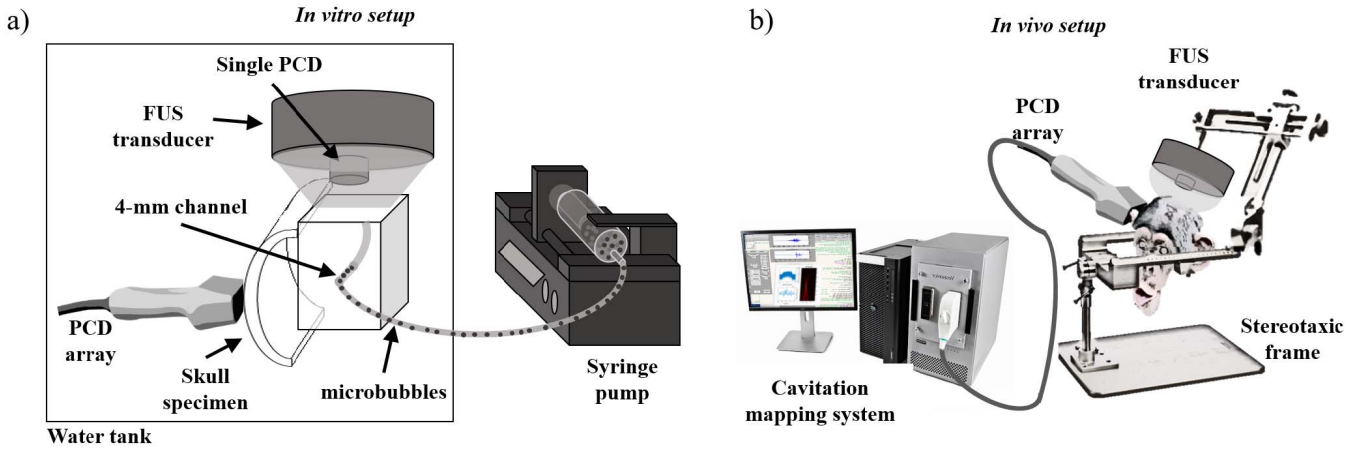


Fig. 3. Experimental setups for (a) *in vitro* skull and phantom and (b) *in vivo* BBB opening in NHP. In the *in vitro* experiment, the FUS transducer was placed on the top of the phantom and orthogonal to the PCD array. The skull was placed between the phantom and the PCD array for assessing the skull effects on the cavitation mapping. In the *in vivo* experiment, the FUS transducer was targeted to the region-of-interest based on the neuronavigation coordinates while the PCD array was placed against the temporal bone window toward the FUS focus.

stereotaxic frame for head fixation [Fig. 3(b)], with targets set at the caudate-putamen and hippocampus using a neuronavigation system (Brainsight Vet System, Rogue Research Inc., Montreal, QC, Canada). The neuronavigation guidance was performed using anatomical T1-weighted magnetic resonance (MR) brain images (3-D turbo field echo sequence,

TR/TE = 11.1/5.1 ms, FA = 8°, and resolution = 0.7 × 0.7 × 0.7 mm^3; Philips 3 Tesla scanner). The animals received in-house manufactured monodisperse microbubbles injected intravenously (2.5×10^8 bubbles/kg) and were sonicated for 2 min (derated peak-negative pressure = 450 kPa, excitation frequency = 0.5 MHz, pulse length = 10 ms, and PRF = 2 Hz). The reported pressures correspond to *in situ* values, following transmission through an NHP or human skull, and were estimated prior to the experiment using a capsule hydrophone (HGL-0200, ±3 dB frequency range: 0.25–40 MHz, electrode aperture: 200 mm; Onda Corp., Sunnyvale, CA, USA). The cavitation activity was monitored in real-time using the same PCD array and single PCD transducer described in the *in vitro* test. The PCD array was aligned with the focal region of the FUS transducer using neuronavigation-guidance through the skull temporal window (a thinner part of the skull serving as an acoustic window). The BBB opening and safety were assessed by MR images acquired 1 h after sonication. The BBB opening was confirmed by comparing T1-weighted contrast-enhanced images (Gd-DTPA-BMA, Omniscan, GE Healthcare, Princeton, NJ, USA; 0.2 mL/kg) acquired before and following the sonication (3-D spoiled gradient echo sequence, TR/TE = 8.5/4.8 ms, FA = 8°, and resolution = 1 × 1 × 1 mm^3). Safety was evaluated with T2-weighted MR images for assessing potential edemas (TR/TE = 3000/80 ms, flip angle or FA = 90°, and resolution = 0.4 × 0.4 × 2 mm^3).

F. Quantification of Acoustic Cavitation Emission Using the Single-Element PCD

Stable cavitation dose (SCD_h), stable cavitation dose with ultra-harmonics (SCD_u), and the inertial cavitation dose (ICD) were calculated following the same methodology of previous studies [58], [59]. Harmonic components with frequency bandwidths of 20 kHz ($n \times f$, where f = 0.5 MHz and $n = 3, 4, 5 \ldots, 10$) were extracted from the frequency spectrum obtained from the PCD signal in Volts. Similarly, ultra-harmonic components ($m/2 \times f$, where f = 0.5 MHz and $m = 5, 7, 9, \ldots, 19$) were extracted using the same frequency bandwidth size. SCD_h was calculated by the root mean square (RMS) of the harmonic components, SCD_u, by the RMS of the ultra-harmonic components and ICD by the RMS of all other components not included in SCD_h and SCD_u, between 1.25 and 5.00 MHz.

G. Benchmarking

The first benchmarking test compared the computational time and sample processing rate of CPU and GPU implementations of the sparse and standard DAS matrix multiplication using the NHP *in vivo* data set (Z = 2000 samples, N = 128 elements, N_z = 2000 axial pixels, and N_x = 128 lateral pixels). All GPU computations were performed using GPU-enabled MATLAB functions. Both GPU and CPU tests were performed using a number of samples N_T from 10 to 1500, except for the CPU standard DAS implementation that was limited to 500 samples as it reached unpractical computational times (several hours). The second benchmarking test compared the computational time with Z = 2000 samples, N = 128 elements, and N_T = 30 for different sizes of FOV, where the lateral map size N_x remained constant equal to 128 pixels, and the axial map size varied: N_z = 50, 100, 200, 500, 1000, and 2000 pixels. The testing routines included only the beamforming processing (Fig. 2) and did not include the sparse matrix construction (Fig. 1). Similarly, the memory allocation for the standard processing was disregarded to allow an adequate comparison of processing time only with both methods. The benchmarking was performed offline in a Dell Precision T7910 workstation (dual-processor Intel Xeon CPU E5-2650 v4 at 2.20 GHz, 128 GB of RAM) equipped with a GPU (NVIDIA Quadro P6000, 24 GB memory, 3840 cores, driver: 392.56) running MS Windows 10 Pro 64-bits and MATLAB 2017b.

III. Results

The effect of the integration time T_i on the GPU sparse-matrix algorithm computational time T_c (including both beamforming time and integration time) and cavitation mapping quality was assessed off-line in phantoms with and without skull specimens. The computational time increased linearly with the integration time as the number of beamformed samples for each cavitation map increased with T_i [Fig. 4(a)]. To achieve real-time monitoring the maximum computational time was limited by the time between pulses ($T_c < 1/\text{PRF}$), which in the case of NHP BBB opening sessions was defined as 0.5 s (for PRF = 2 Hz). A maximum T_i of 1.44 μs was found in order to achieve a real-time cavitation mapping with N_T equaling 30 samples. In addition to that, the maximal intensity in the mapping plateaued at approximately 62.5 μs (1300 samples). The maximum intensity for each integration time was defined as the pixel intensity with the highest value in each reconstructed passive map. The −6 dB cavitation region size defined in the map was quantified and found to increase with T_i with a transient formation of discrete spots of cavitation activity during the first 20 μs [Fig. 4(b)]. Then, the cavitation region size decreased, possibly due to the destruction of resonant microbubbles at the periphery of the focus, reaching a steady-state spatial distribution at around 62.5 μs.

The cavitation detectability determined by system sensitivity was then tested in phantoms using 62.5-μs integration time for pressure levels ranging from 150 to 600 kPa (Fig. 5). Cavitation maps without the skull showed localized cavitation distributions at all pressure levels [Fig. 5(a)]. Acquisitions with skull samples presented a threshold for cavitation detection at 300 and 450 kPa for NHP skull [Fig. 5(b)] and human skull [Fig. 5(c)], respectively. The cavitation activity was spatially distorted in the presence of both skulls, forming an elongated pattern as a result of the beamforming degradation caused by the skull scattering. The pressure thresholds identified here are in the range used for BBB opening in NHP.

Following the *in vitro* experiments, the *in vivo* experiments were performed in NHP during the BBB opening sessions. The PCD array was placed on the temporal window aiming at the FUS targeted area, with the PAM plane covering a lateral cross section of the FUS focus [Fig. 6(a)]. The BBB opening

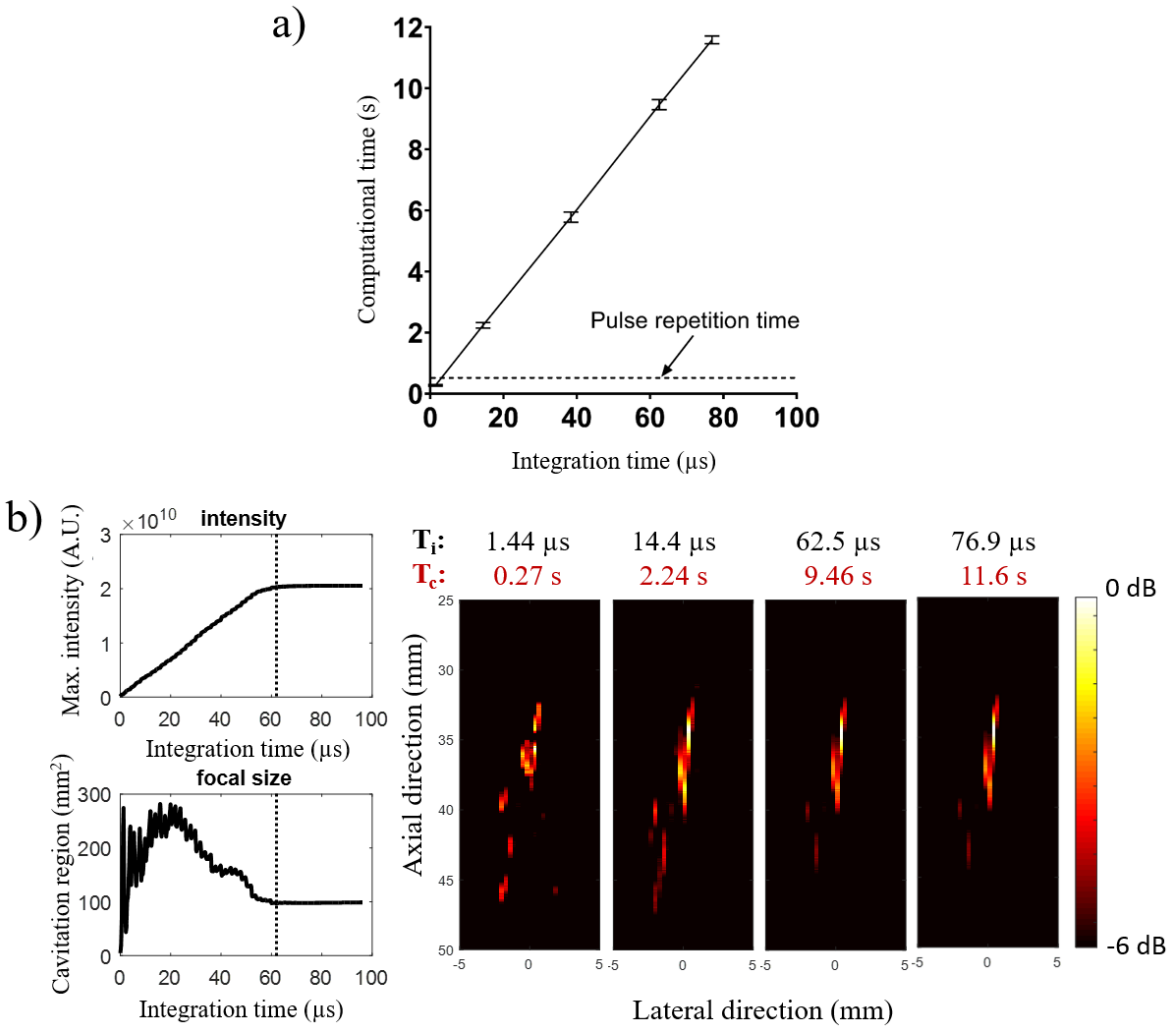


Fig. 4. Effect of integration time on computational time and cavitation mapping characteristics. (a) Computational time T_c increased linearly with the integration time T_i, which limited the integration time to a maximum of 1.44 μs for PRF = 2 Hz (0.5 s of pulse repetition period). (b) T_i also affected the mapping quality with lower values providing maps with discrete cavitation spots out of focus and values higher than 62.5 μs reaching a steady state of cavitation map. This representative case was performed at 450 kPa. Computational times T_c refers to the reconstruction of large acoustic maps (2000 axial pixels × 128 lateral pixels).

and monitoring result at 450 kPa obtained from NHP 1 is shown in Fig. 6(a). The frequency spectra from the beamformed signals at the location of maximum image intensity acquired during control and acquisitions with microbubbles are shown in Fig. 6(b). The cavitation levels obtained from the single-element PCD acquisition following spectral filtering showed a substantial increase of both stable and inertial cavitation once microbubbles perfused the brain [Fig. 6(c)]. The cavitation intensity [Fig. 6(d)] obtained from consecutive cavitation maps [Fig. 6(e)] was qualitatively consistent with the cavitation activity observed with the single-element PCD. The single-element PCD traces are presented with independent components cavitation (harmonic, ultra-harmonic, and inertial), whereas the cavitation activity detected by the PCD array shows the total activity without spectral filtering.

The results from the second test performed at 450 kPa in the NHP 2 are shown in Fig. 7. In this case, only stable cavitation was observed from data acquired with the single-element PCD [Fig. 7(c)]. The majority of the observed stable cavitation was harmonic-based, with only a few pulses having ultra-harmonic dose higher than the baseline. The frequency spectra obtained with the PCD array at the location of maximum image intensity are shown in Fig. 7(b). As shown in the previous studies, the skull attenuation is highly variable across different skull locations and across animals due to bone thickness variation and different ratios of cortical to trabecular bone [22], [60]. The variation of the skull attenuation contributed to the differences in the signal components using the single-element and PCD array [Fig. 7(c) and (d)]. In addition, the single-element PCD presents a higher sensitivity and much broader frequency bandwidth. The single-element PCD had a higher sampling frequency (50 MHz versus 20.8 MHz) than the PCD array, thereby providing spectra at higher frequencies without aliasing. These differences highlight the importance of multiple cavitation detectors for safety redundancy. The cavitation maps in the logarithmic intensity scale (in decibel)

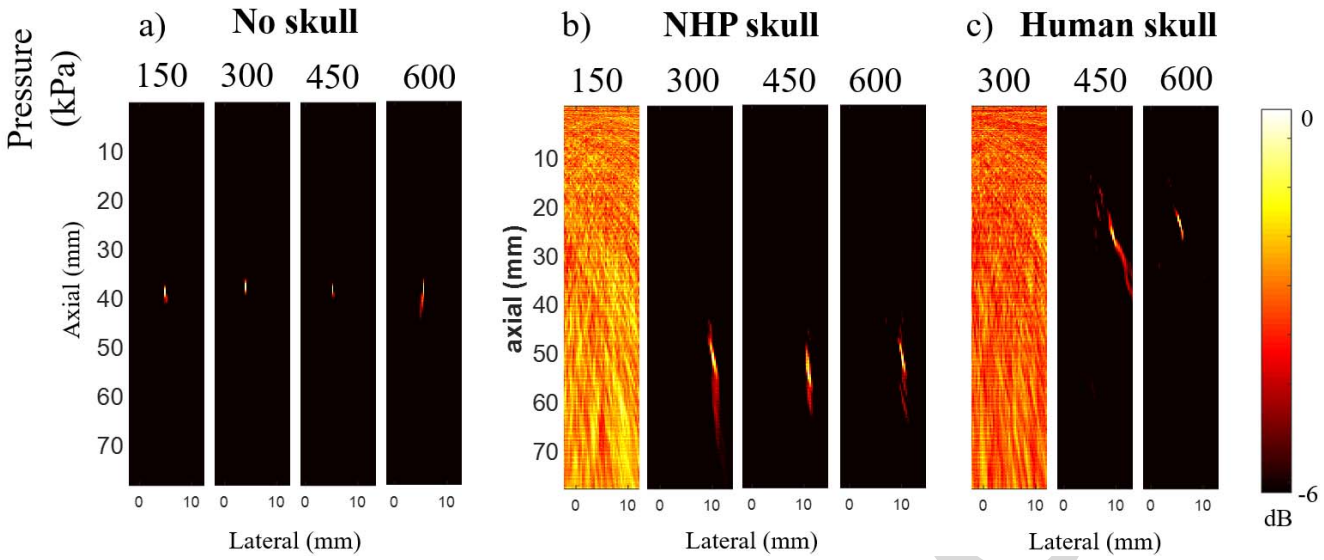


Fig. 5. Cavitation mapping sensitivity through primate skull. Cavitation maps using 62.5-μs integration time at variable pressure levels were acquired (a) without a skull, (b) with NHP skull, and (c) with human skull between the PCD array and the phantom.

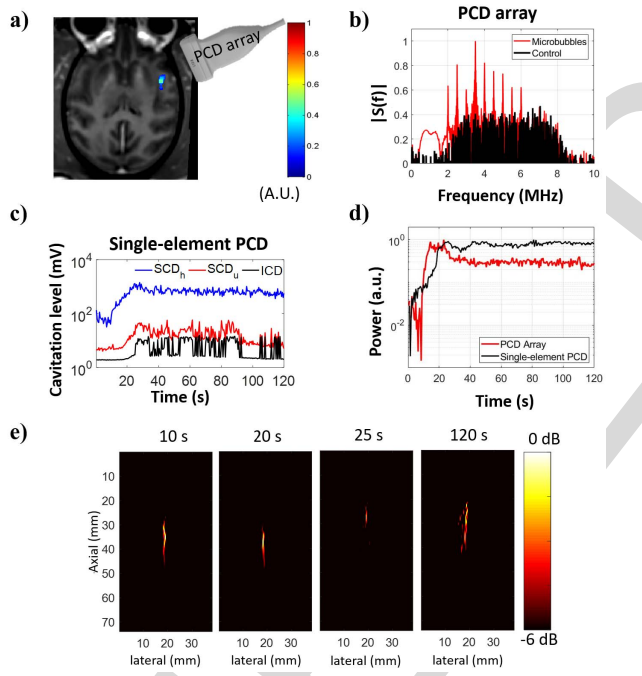


Fig. 6. Cavitation activity recorded during BBB opening in NHP 1. (a) BBB opening (in color) induced by sonication at 450 kPa revealed in contrast-enhanced T1-weighted MR image. (b) Frequency spectra obtained from the beamformed signal at the location of maximum image intensity. (c) Cavitation activity evaluated with a single-element PCD transducer (cavitation dose). (d) Normalized power detected with a PCD array positioned at the temporal window and a single-element PCD transducer co-aligned with the FUS transducer. Power was defined as the sum of squared beamformed data using 1.44-μs integration time for the PCD array and the entire signal duration for the single-element PCD transducer. (e) Spatial distribution of the cavitation activity shown in cavitation maps (−6 dB) for different samples.

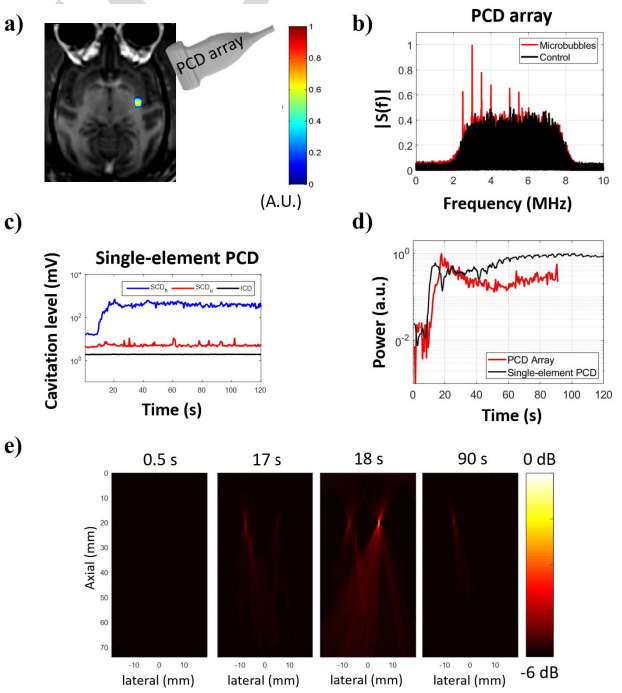


Fig. 7. Cavitation activity recorded during BBB opening in NHP 2. (a) BBB opening (in color) induced by sonication at 450 kPa revealed in contrast-enhanced T1-weighted MR image. (b) Frequency spectra obtained from the beamformed signal at the location of maximum image intensity. (c) Cavitation dose detected using a single-element PCD transducer indicating only stable cavitation throughout the sonication duration. (d) Normalized power detected with a PCD array positioned at the temporal window and a single-element PCD transducer co-aligned with the FUS transducer using 1.44-μs integration time. (e) Reconstructed cavitation maps (−6 dB) for different samples.

were used for monitoring both the spatial location and intensity of cavitation over time [Fig. 7(e)]. Interestingly, cavitation events were detected in two locations, possibly caused by cavitation in the neighboring larger vessels or caused by the sidelobes of the FUS beam.

The benchmarking revealed that sparse matrix operation improved both GPU and CPU performance. The computational

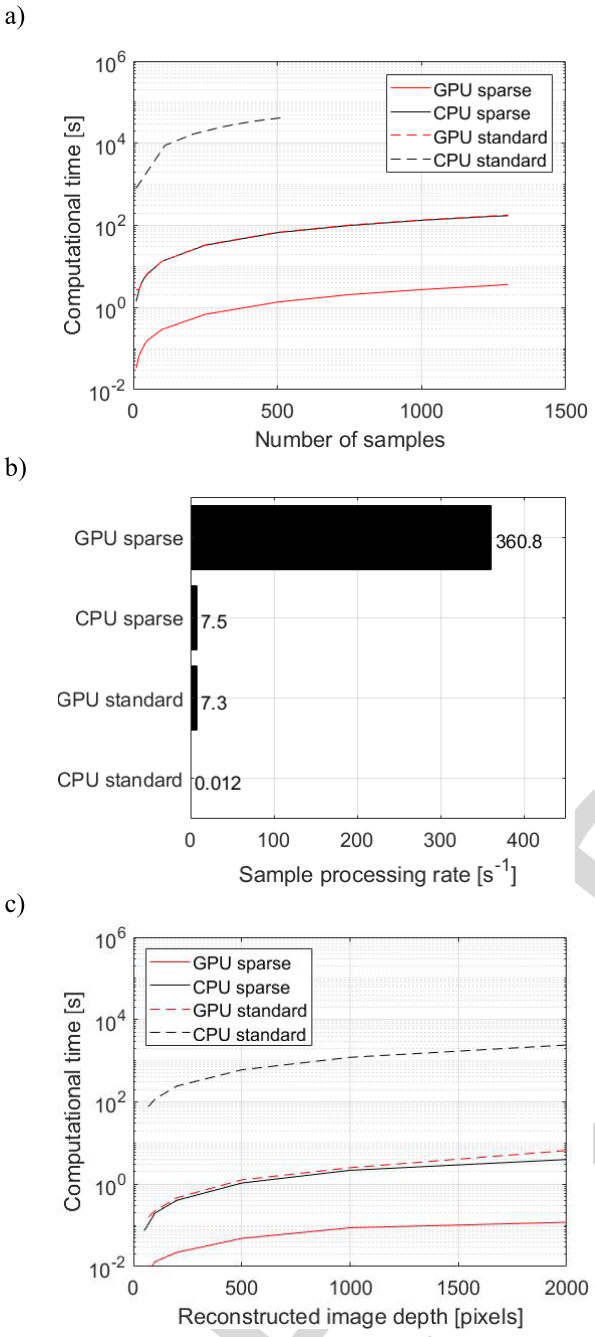


Fig. 8. Benchmarking for sparse matrix operation. (a) Computational time of CPU and GPU implementations for the sparse matrix multiplication and standard DAS using the NHP *in vivo* data set for integrated sample N_T varying from 10 to 1500 (Z: 2000 samples, N: 128 elements, N_x: 128 lateral pixels, and N_z: 2000 axial pixels). (b) Sample processing rate at the same conditions. (c) Computational time with $N_T = 30$ for different sizes of FOV, where the lateral map size N_x remained constant equal to 128 pixels and the axial map size varied: $N_z = 50$, 100, 200, 500, 1000, and 2000 pixels. The benchmarking was performed offline in a Dell Precision T7910 workstation (dual processor Intel Xeon CPU E5-2650 v4 at 2.20 GHz, 128 GB of RAM) equipped with a GPU (NVIDIA Quadro P6000, 24 GB memory, 3840 cores, driver: 392.56) running MS Windows 10 Pro 64 bits and MATLAB 2017b.

time had a linear relation with the number of samples, and it is presented in the log scale as the CPU standard DAS implementation resulted in computational times four orders of magnitude higher than the GPU sparse implementation [Fig. 8(a)]. In GPU, the sparse matrix performed approximately 50 times faster than the standard DAS matrix, whereas, in CPU, the same comparison resulted in approximately 600 times difference. Interestingly, the CPU sparse and GPU standard DAS implementations presented similar results with sample processing rates of 7.3 and 7.5 s^{-1}, respectively [Fig. 8(b)]. On the other side, the GPU standard DAS implementation presented a much higher average sample processing rate of 360.8 s^{-1}. Similarly, for a fixed number of samples $N_T = 30$, the sparse matrix implementation reduced the computational time for variable FOV sizes [Fig. 8(c)]. The GPU sparse matrix implementation decreased the computational time by a minimum of 21.9 times (FOV: 50 $\times$ 128 pixels) in comparison with GPU standard matrix implementation and a maximum of 55.3 times (FOV: 2000 $\times$ 128 pixels) with an average of 28.5 times across different FOVs. The sparse matrix performed on CPU was an average of 617 times faster than the CPU standard implementation and 1.33 times faster than the GPU standard implementation. The processing times required per PAM pixel for FOV varying from 50 to 2000 pixels axially $\times$ 128 pixels laterally using 30 samples were on average 0.03 $\pm$ 0.01, 0.51 $\pm$ 0.01, 0.67 $\pm$ 0.09, and 311.71 $\pm$ 3.94 μs/pixel/sample for GPU sparse, CPU sparse, GPU standard, and CPU standard implementations, respectively. For both CPU and GPU implementations, the sparse matrix was loaded in the memory prior to the time testing routines. Similarly, for the standard DAS method, the time for memory allocation was disregarded. The loading time was 6.86 $\pm$ 0.13 s and the CPU–GPU transfer time was 0.79 $\pm$ 0.06 s.

IV. Discussion

In this study, we demonstrated the implementation of sparse matrix beamforming and TEAs on a GPU for real-time transcranial cavitation mapping. The system was tested *in vitro* using human and NHP skull specimens, which allowed to investigate the computational time for real-time cavitation mapping, and *in vivo* during ultrasound-mediated BBB opening sessions in NHP, which allowed to test the setup and algorithm in a close to clinical setup condition.

The summation over the integration time used in the TEA algorithm enhanced the continuously scattered signal from microbubbles by suppressing the background noise, which affected the mapping quality and homogeneity of cavitation distribution [61], [62]. Noisier maps (discrete cavitation spots outside the transducer focus potentially associated with very transient bubble activity) were found *in vitro* for short integration times, while a steady cavitation distribution could be achieved with long integration time duration (>62.5 μs). Despite the higher SNR achieved with a high number of beamformed samples, the computational load and the number of frames were limited by the total integration time to lie within the pulse repetition period. For the PRF used in the test *in vivo* (2 Hz), 30 samples could be reconstructed for an integration time of 1.44 μs. The increased number of pixels in the axial direction (i.e., 2000) was chosen based on active imaging parameters; however, the pixel size of 37 μm is significantly smaller than the nominal passive acoustic mapping resolution at these imaging depths. The nominal axial resolution of PAM would be equal to 2.1 mm, assuming an

imaging depth of 40 mm, an aperture size of 38.4 mm, and a mean receive wavelength of 0.28 mm [63]. As demonstrated by the benchmarking, the sparse matrix density ($N_x \times N_z$ by $N \times Z$) and the RF data set density ($N \times Z$ by N_T) [Fig. 2(a)] can be modified, which ultimately changes the computational time at a rate of 0.03 ± 0.01 μs/pixel/sample. Therefore, a reduced number of pixels can accelerate the processing, which allows a longer integration time. Future work will involve cavitation maps with a larger axial pixel size, to accelerate computation times and enable real-time mapping with the integration of longer data sets, while preserving spatial information. As previously reported by Acconcia *et al.* [64], the microseconds time scale can potentially provide insights into bubble dynamics, such as the rapid bubble cloud evolution and its stochastic nature, as opposed to long integration times. Furthermore, the majority of cavitation activity is constrained within the first hundreds of microseconds of an ms-long therapeutic pulse. Previous *in vitro* work has shown that 80% of the total cavitation energy is emitted within 200 μs (or 0.2% of the total pulse length) during therapeutic ultrasound exposure [65], [66]. This is likely due to the destruction of resonant microbubbles at the beginning of the therapeutic pulses, with smaller nonresonant microbubbles emitting lower cavitation energy for the remainder of the pulse. Future studies will correlate time exposure with bubble dynamics observed with high speed videography [67]. The pressure thresholds for the cavitation detection through human and NHP parietal bone specimens using the PCD array were in the range of pressure employed in previous studies [25], [58], [59], [68]. These results demonstrate the capability of the system to map spatial and temporal microbubble activity in real-time.

The benchmarking results revealed that the sparse matrix operation can decrease considerably the computational time in both CPU and GPU. For the computer configuration tested here, the CPU sparse implementation performed very similar to the GPU standard DAS implementation. Despite differences that other computer configurations may present (i.e., less powerful CPU processors with more powerful GPU), the sparse matrix operation is demonstrated to be a feasible solution for decreasing the computational time of operations with dense data sets, which enables either larger FOV or larger data sets relevant to therapeutic applications. An interesting characteristic of the sparse matrix operation is that it can be applied regardless of the number of processed samples, thus the integration time can be adapted easily to result in high image quality based on the trade-off of integration time and real-time visualization. It is important to note that similar operation is not practical with a fully sampled reconstruction matrix as for the data set used here matrix ($256\,000 \times 256\,000$), the fully sampled matrix would require 488.3 GB of computer memory as opposed to sparse representation that requires only 1.05 GB. On the other hand, the sparse matrix-based beamforming presents limitations such as the need to construct the sparse matrix [(3)–(5)] offline and several hours prior to its use. This method requires time-consuming iterative calculations and memory allocation to allow the element indentation. This is partially resolved as only a single matrix construction at the highest sampling is needed. Then, the matrix can be downsampled to adjust to any reduced FOV or channel data size. Once the sparse matrix size is adjusted, the multiplication of the sparse matrix and channel data can be performed, which will result in the beamformed data. The allocation of the pre-constructed sparse matrix in the computer or GPU memory is performed only once and it takes a few seconds, which will depend on the computer performance and sparse matrix size.

Next, we demonstrated an *in vivo* test of our system during FUS-mediated BBB opening session in NHP. The skull presents high variability in the thickness, variable proportion of cortical, and trabecular bone distribution across skull regions, and, subsequently, high variability of the ultrasound attenuation [60]. Therefore, it is important to monitor cavitation activity with as low attenuation as possible, especially since high frequency emissions from microbubbles are more heavily attenuated through the skull. Cavitation mapping was, therefore, performed through the thin temporal bone, which presents lower attenuation than other skull bones. This experimental configuration was shown to be viable at mapping cavitation activity and demonstrated the feasibility of GPU-accelerated sparse matrix computing in a close to clinical setup. The system was qualitatively compared against single-element PCD acquisitions, which presented the same trend for the cavitation activity. PCD with both single-element and multielement transducers was performed to illustrate the qualitative similarities in the temporal evolution of the detected signals [Figs. 6(d) and 7(d)]. However, we have used the normalized values, because a direct quantitative comparison is not possible, due to different receiving center frequencies, sensitivity patterns, and sampling frequencies. In the foreseen clinical use, device limitations can be ameliorated by using multiple safety monitoring redundancies, which in this case both single-element and array transducers can detect independently potentially high-risk cavitation activity. Future implementations will include simulations based on CT-scans [35], [37], [69] with the co-registration of MR images from the neuronavigator and simulated acoustic beam profiles.

Beamforming degradation caused by the diffraction pattern of the receiving array and interference caused by multiple bubbles emitting acoustic signals at the same time generated spatial distortions of cavitation activity, which resulted in an elongated pattern of cavitation activity. Differences in the beam shift can be explained due to differences in the skull curvatures of human and NHP specimens. As previously reported [32], [61], [69], potential focal shifts may be present, but were not studied here. Differences in the human and NHP specimen size forced different positioning of the imaging array in relation to the therapeutic transducer. Thus, the location of focus was not consistent across acquisitions. Nevertheless, the focal distortion is an example of problems encountered during treatment that shows the need of real-time spatial monitoring of cavitation.

Frequency-domain approaches can significantly accelerate processing even without a GPU, due to the spectral filtering of narrowband harmonics or ultraharmonics [35], [36]. However, the quality and speed of such approaches depends on the pulse length. Shortening the pulse length leads to progressively wider harmonic and ultra-harmonic peaks, thereby

increasing the integration bandwidth and computational load. Furthermore, discrimination of different cavitation modes, e.g., inertial versus non-inertial cavitation, becomes increasingly more difficult with shorter pulse lengths. Therefore, frequency-domain algorithms need to be modified for different excitation sequences, especially for the examined frequency windows. Finally, to calculate the total acoustic energy emitted by the exposed microbubbles, one needs to integrate throughout the frequency domain, which negates the need for frequency-based analysis.

On the other hand, time-domain approaches can be very slow when examining long RF signals with a traditional CPU approach [Fig. 8(a)]. Time-domain DAS is a simple process inducing a relatively low computational load and provides direct information about the total cavitation energy produced within the focus, regardless of the pulse length or cavitation mode. Cavitation energy has been previously correlated with the induced bioeffect, such as the drug delivery efficiency [70]. The same algorithm can be applied without modification using any excitation sequences, ranging from μs-long pulse sequences for BBB opening [71] to s-long sequences for HIFU ablation [30]. Our study describes a method for accelerating time-domain PAM for both CPU- and GPU-based systems (Fig. 8) across a broad spectrum of ultrasound therapies. Sparse-matrix multiplication provides a general method to accelerate PAM processing at any pulse length [Fig. 8(a)] or map size [Fig. 8(c)].

The sparse matrix implementation demonstrated here employed a simple DAS algorithm that allowed for accelerated image acquisitions. However, as previously described by Haworth *et al.* [36], the axial resolution using this algorithm can be ten times worse than lateral resolution when using small-aperture 1-D receiver arrays. When large-aperture 2-D receiver arrays are employed the axial dimension of the point-spread function obtained via DAS beamforming is approximately twice that of its lateral counterpart [61], [72], [73]. As we demonstrated in NHP, this can be partially overcome by changing the orientation of the passive array at an acute angle using the temporal window so the mapping can be provided with a compromise of resolution between the axial and lateral orientations (worst case in coaxial orientation with the therapy transducer). In addition, introduction of the skull may have caused shifts in the position of cavitation activity in reconstructed images. Yet, the algorithm tested in this study demonstrates the capability of the GPU-accelerated sparse matrix operation using large RF data sets. Despite the higher spatial resolution achieved with more sophisticated algorithms such as the robust Capon beamforming [74], [75], its computational load is still challenging for GPU-processing. We recognize that including a comparison against other cavitation mapping algorithms would have been desirable; however, limited key information and expertise in alternative methods would yield misrepresentative conclusions. Coded excitation can also be implemented to improve the confinement of microbubble activity [76] and the decoding on receive could enable higher SNR [77]. Additionally, implementation of a closed-loop feedback control for cavitation mapping may be beneficial for online treatment optimization [78]. This would provide a more reliable control of the microbubble activity from regions inside the brain while avoiding artifact from other sources such as acoustic coupling media or regions outside the brain [22]. In our future work, L7-4 will be replaced with a lower frequency array, to facilitate cavitation mapping through the human skull and reduce aberrations occurring at higher frequencies.

Finally, the sparse matrix implementation in GPU is a feasible method for accelerating image formation. This method can enable enlarging the FOV that could potentially help identify in real time off-target cavitation activity. In the context of this study, real-time imaging was enabled only for extremely short integration times (i.e., 1.44-μs duration or 30 samples). Future work will involve generation of maps with a reduced number of pixels but larger pixel sizes in the axial direction, in order to accelerate beamforming of longer data sets. Such a reduction of the axial pixel number by—for example—50-fold would lead to an equivalent increase of the permissible integration time by 50-fold. Together with the neuronavigation system [25], the therapeutic transducer could be realigned at normal incidence angles [79] during sonication or the sonication could be halted if substantial beam aberrations are observed.

V. Conclusion

Passive acoustic mapping has a great potential in clinical cavitation-based FUS applications especially for monitoring and guiding the treatments. A detailed implementation of sparse matrix beamforming on a GPU for cavitation mapping is demonstrated here as a method to accelerate processing. We demonstrate with *in vitro* and *in vivo* tests that the GPU-based sparse matrix method can accelerate passive acoustic mapping compared to standard GPU or CPU processing, and allow real-time processing of large maps (e.g., 2000 axial pixels $\times$ 128 lateral pixels) with limited integration times (e.g., 30 samples). This methodology was proven efficient for both CPU and GPU implementations. Moreover, cavitation mapping through the human skull bone showed the feasibility to use this technique in clinical applications. Finally, the real-time capability together with the neuronavigation system enables the operator to correct or halt sonications in case substantial aberrations are observed.

References

[1] M. Hammarlund-Udenaes, E. C. M. de Lange, and R. G. Thorne, Eds., *Drug Delivery to the Brain*, vol. 10. New York, NY, USA: Springer, 2014.

[2] K. Hynynen, N. McDannold, N. Vykhodtseva, and F. A. Jolesz, "Non-invasive MR imaging–guided focal opening of the blood-brain barrier in rabbits," *Radiology*, vol. 220, no. 3, pp. 640–646, Sep. 2001.

[3] M. Aryal, K. Fischer, C. Gentile, S. Gitto, Y.-Z. Zhang, and N. McDannold, "Effects on P-Glycoprotein expression after blood-brain barrier disruption using focused ultrasound and microbubbles," *PLoS ONE*, vol. 12, no. 1, Jan. 2017, Art. no. e0166061.

[4] H. Cho *et al.*, "Localized down-regulation of P-glycoprotein by focused ultrasound and microbubbles induced blood-brain barrier disruption in rat brain," *Sci. Rep.*, vol. 6, no. 1, p. 31201, Nov. 2016.

[5] Z. Zhang *et al.*, "Low intensity ultrasound promotes the sensitivity of rat brain glioma to doxorubicin by down-regulating the expressions of P-Glucoprotein and multidrug resistance protein 1 *in vitro* and *in vivo*," *PLoS ONE*, vol. 8, no. 8, Aug. 2013, Art. no. e70685.

[6] M. Kinoshita, N. McDannold, F. A. Jolesz, and K. Hynynen, "Noninvasive localized delivery of Herceptin to the mouse brain by MRI-guided focused ultrasound-induced blood-brain barrier disruption.," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 31, pp. 23–11719, 2006.

[7] H.-L. Liu *et al.*, "Blood-brain barrier disruption with focused ultrasound enhances delivery of chemotherapeutic drugs for glioblastoma treatment," *Radiology*, vol. 255, no. 2, pp. 415–425, 2010.

[8] L. H. Treat, N. McDannold, Y. Zhang, N. Vykhodtseva, and K. Hynynen, "Improved anti-tumor effect of liposomal doxorubicin after targeted blood-brain barrier disruption by MRI-guided focused ultrasound in rat glioma," *Ultrasound Med. Biol.*, vol. 38, no. 10, pp. 1716–1725, Oct. 2012.

[9] J. F. Jordão *et al.*, "Antibodies targeted to the brain with image-guided focused ultrasound reduces amyloid-β plaque load in the TgCRND8 mouse model of Alzheimer's disease," *PLoS ONE*, vol. 5, no. 5, pp. 4–11, 2010.

[10] S. B. Raymond, L. H. Treat, J. D. Dewey, N. J. McDannold, K. Hynynen, and B. J. Bacskai, "Ultrasound enhanced delivery of molecular imaging and therapeutic agents in Alzheimer's disease mouse models," *PLoS ONE*, vol. 3, no. 5, pp. 1–7, 2008.

[11] B. Baseri *et al.*, "Activation of signaling pathways following localized delivery of systemically administered neurotrophic factors across the blood–brain barrier using focused ultrasound and microbubbles," *Phys. Med. Biol.*, vol. 57, no. 7, pp. N65–N81, Apr. 2012.

[12] H. Chen, G. Z. X. Yang, H. Getachew, C. Acosta, C. Sierra Sánchez, and E. E. Konofagou, "Focused ultrasound-enhanced intranasal brain delivery of brain-derived neurotrophic factor," *Sci. Rep.*, vol. 6, no. 1, p. 28599, Sep. 2016.

[13] R. Ji *et al.*, "Focused ultrasound enhanced intranasal delivery of brain derived neurotrophic factor produces neurorestorative effects in a Parkinson's disease mouse model," *Sci. Rep.*, vol. 9, no. 1, p. 19402, Dec. 2019.

[14] A. Alonso *et al.*, "Focal delivery of AAV2/1-transgenes into the rat brain by localized ultrasound-induced BBB opening," *Mol. Therapy-Nucleic Acids*, vol. 2, no. 1, p. e73, Jan. 2013.

[15] S. Wang, O. O. Olumolade, T. Sun, G. Samiotaki, and E. E. Konofagou, "Noninvasive, neuron-specific gene therapy can be facilitated by focused ultrasound and recombinant adeno-associated virus," *Gene Therapy*, vol. 22, no. 1, pp. 104–110, Jan. 2015.

[16] A. Burgess, C. A. Ayala-Grosso, M. Ganguly, J. F. Jordão, I. Aubert, and K. Hynynen, "Targeted delivery of neural stem cells to the brain using MRI-guided focused ultrasound to disrupt the blood-brain barrier," *PLoS ONE*, vol. 6, no. 11, Nov. 2011, Art. no. e27877.

[17] Y. Meng *et al.*, "Blood-brain barrier opening in Alzheimer's disease using MR-guided focused ultrasound," *Neurosurgery*, vol. 66, no. 1, p. 2336, Sep. 2019.

[18] T. Mainprize *et al.*, "Blood-brain barrier opening in primary brain tumors with non-invasive MR-guided focused ultrasound: A clinical safety and feasibility study," *Sci. Rep.*, vol. 9, no. 1, p. 321, Dec. 2019.

[19] M. A. O'Reilly and K. Hynynen, "Blood-brain barrier: Real-time feedback-controlled focused ultrasound disruption by using an acoustic Emissions–based controller," *Radiology*, vol. 263, no. 1, pp. 96–106, Apr. 2012.

[20] C.-H. Tsai, J.-W. Zhang, Y.-Y. Liao, and H.-L. Liu, "Real-time monitoring of focused ultrasound blood-brain barrier opening via subharmonic acoustic emission detection: Implementation of confocal dual-frequency piezoelectric transducers," *Phys. Med. Biol.*, vol. 61, no. 7, pp. 2926–2946, Apr. 2016.

[21] T. Sun *et al.*, "Closed-loop control of targeted ultrasound drug delivery across the blood–brain/tumor barriers in a rat glioma model," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 48, pp. E10281–E10290, Nov. 2017.

[22] H. A. Kamimura *et al.*, "Feedback control of microbubble cavitation for ultrasound-mediated blood–brain barrier disruption in non-human primates under magnetic resonance guidance," *J. Cereb. Blood Flow Metab.*, vol. 39, no. 7, pp. 1191–1203, Jan. 2018.

[23] N. McDannold, C. D. Arvanitis, N. Vykhodtseva, and M. S. Livingstone, "Temporary disruption of the blood-brain barrier by use of ultrasound and microbubbles: Safety and efficacy evaluation in rhesus macaques," *Cancer Res.*, vol. 72, no. 14, pp. 3652–3663, Jul. 2012.

[24] Y.-S. Tung, J. J. Choi, B. Baseri, and E. E. Konofagou, "Identifying the inertial cavitation threshold and skull effects in a vessel phantom using focused ultrasound and microbubbles," *Ultrasound Med. Biol.*, vol. 36, no. 5, pp. 840–852, May 2010.

[25] S.-Y. Wu *et al.*, "Efficient blood-brain barrier opening in primates with neuronavigation-guided ultrasound and real-time acoustic mapping," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, Dec. 2018.

[26] V. A. Salgaonkar, S. Datta, C. K. Holland, and T. D. Mast, "Passive cavitation imaging with ultrasound arrays," *J. Acoust. Soc. Amer.*, vol. 126, no. 6, pp. 3071–3083, Dec. 2009.

[27] M. Gyongy and C.-C. Coussios, "Passive spatial mapping of inertial cavitation during HIFU exposure," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 1, pp. 48–56, Jan. 2010.

[28] M. D. Gray, E. Lyka, and C. C. Coussios, "Diffraction effects and compensation in passive acoustic mapping," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 2, pp. 258–268, Feb. 2018.

[29] K. J. Haworth *et al.*, "Passive imaging with pulsed ultrasound insonations," *J. Acoust. Soc. Amer.*, vol. 132, no. 1, pp. 544–553, Jul. 2012.

[30] C. R. Jensen, R. W. Ritchie, M. Gyöngy, J. R. T. Collin, T. Leslie, and C.-C. Coussios, "Spatiotemporal monitoring of high-intensity focused ultrasound therapy with passive acoustic mapping," *Radiology*, vol. 262, no. 1, pp. 252–261, Jan. 2012.

[31] K. J. Haworth, V. A. Salgaonkar, N. M. Corregan, C. K. Holland, and T. D. Mast, "Using passive cavitation images to classify high-intensity focused ultrasound lesions," *Ultrasound Med. Biol.*, vol. 41, no. 9, pp. 2420–2434, Sep. 2015.

[32] C. D. Arvanitis, G. T. Clement, and N. McDannold, "Transcranial assessment and visualization of acoustic cavitation: Modeling and experimental validation," *IEEE Trans. Med. Imag.*, vol. 34, no. 6, pp. 1270–1281, Jun. 2015.

[33] C. Coviello *et al.*, "Passive acoustic mapping utilizing optimal beamforming in ultrasound therapy monitoring," *J. Acoust. Soc. Amer.*, vol. 137, no. 5, pp. 2573–2585, May 2015.

[34] E. Lyka, C. Coviello, R. Kozick, and C.-C. Coussios, "Sum-of-harmonics method for improved narrowband and broadband signal quantification during passive monitoring of ultrasound therapies," *J. Acoust. Soc. Am.*, vol. 140, no. 1, p. 741, Jul. 2016.

[35] C. D. Arvanitis, C. Crake, N. McDannold, and G. T. Clement, "Passive acoustic mapping with the angular spectrum method," *IEEE Trans. Med. Imag.*, vol. 36, no. 4, pp. 983–993, Apr. 2017.

[36] K. J. Haworth, K. B. Bader, K. T. Rich, C. K. Holland, and T. D. Mast, "Quantitative frequency-domain passive cavitation imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 64, no. 1, pp. 177–191, Jan. 2017.

[37] R. M. Jones, M. A. O'Reilly, and K. Hynynen, "Experimental demonstration of passive acoustic imaging in the human skull cavity using CT-based aberration corrections," *Med. Phys.*, vol. 42, no. 7, pp. 4385–4400, Jun. 2015.

[38] J. Collin, C. Coviello, E. Lyka, T. Leslie, and C. Coussios, "Real-time three-dimensional passive cavitation detection for clinical high intensity focused ultrasound systems," *J. Acoust. Soc. Amer.*, vol. 133, no. 5, p. 3263, 2013.

[39] E. Lyka, C. M. Coviello, C. Paverd, M. D. Gray, and C. C. Coussios, "Passive acoustic mapping using data-adaptive beamforming based on higher order statistics," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2582–2592, Dec. 2018.

[40] R. M. Jones, L. Deng, K. Leung, D. Mcmahon, M. A. O'Reilly, and K. Hynynen, "Three-dimensional transcranial microbubble imaging for guiding volumetric ultrasound-mediated blood-brain barrier opening," *Theranostics*, vol. 8, no. 11, pp. 2909–2926, 2018.

[41] L. Zhuo and V. K. Prasanna, "Sparse matrix-vector multiplication on FPGAs," in *Proc. ACM/SIGDA 13th Int. Symp. Field-Programmable Gate Arrays (FPGA)*, 2005, p. 63.

[42] K. K. Matam and K. Kothapalli, "Accelerating sparse matrix vector multiplication in iterative methods using GPU," in *Proc. Int. Conf. Parallel Process.*, Sep. 2011, pp. 612–621.

[43] R. Fang, T. Chen, D. Metaxas, P. Sanelli, and S. Zhang, "Sparsity techniques in medical imaging," *Computerized Med. Imag. Graph.*, vol. 46, p. 1, Dec. 2015.

[44] H. Deshpande, P. Maurel, and C. Barillot, "Classification of multiple sclerosis lesions using adaptive dictionary learning," *Computerized Med. Imag. Graph.*, vol. 46, pp. 2–10, Dec. 2015.

[45] A. Neubert, J. Fripp, C. Engstrom, D. Schwarz, M.-A. Weber, and S. Crozier, "Statistical shape model reconstruction with sparse anomalous deformations: Application to intervertebral disc herniation," *Computerized Med. Imag. Graph.*, vol. 46, pp. 11–19, Dec. 2015.

[46] E. Belilovsky *et al.*, "Predictive sparse modeling of fMRI data for improved classification, regression, and visualization using the k-support norm," *Computerized Med. Imag. Graph.*, vol. 46, pp. 40–46, Dec. 2015.

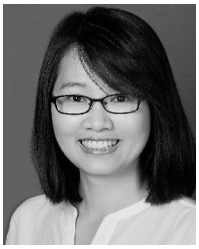
[47] J. Xu *et al.*, "Sparse non-negative matrix factorization (SNMF) based color unmixing for breast histopathological image analysis," *Computerized Med. Imag. Graph.*, vol. 46, pp. 20–29, Dec. 2015.

[48] T. Chen and C. Srinivas, "Group sparsity model for stain unmixing in brightfield multiplex immunohistochemistry images," *Computerized Med. Imag. Graph.*, vol. 46, pp. 30–39, Dec. 2015.
[49] J. Zhou *et al.*, "Automated compromised right lung segmentation method using a robust atlas-based active volume model with sparse shape composition prior in CT," *Computerized Med. Imag. Graph.*, vol. 46, pp. 47–55, Dec. 2015.
[50] R. Fang, H. Jiang, and J. Huang, "Tissue-specific sparse deconvolution for brain CT perfusion," *Computerized Med. Imag. Graph.*, vol. 46, pp. 64–72, Dec. 2015.
[51] G. Y. Hou *et al.*, "Sparse matrix beamforming and image reconstruction for 2-D HIFU monitoring using harmonic motion imaging for focused ultrasound (HMIFU) with *in vitro* validation," *IEEE Trans. Med. Imag.*, vol. 33, no. 11, pp. 2107–2117, Nov. 2014.
[52] J. Grondin, T. Payen, S. Wang, and E. E. Konofagou, "Real-time monitoring of high intensity focused ultrasound (HIFU) ablation of and *in vitro* and canine livers using harmonic motion imaging for focused Uultrasound (HMIFU)," *J. Vis. Exp.*, vol. 105, Nov. 2015, Art. no. e53050.
[53] C. Schretter, S. Bundervoet, D. Blinder, A. Dooms, J. D'hooge, and P. Schelkens, "Ultrasound imaging from sparse RF samples using system point spread functions," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 3, pp. 316–326, Mar. 2018.
[54] E. Roux, F. Varray, L. Petrusca, C. Cachard, P. Tortoli, and H. Liebgott, "Experimental 3-D ultrasound imaging with 2-D sparse arrays using focused and diverging waves," *Sci. Rep.*, vol. 8, no. 1, pp. 1–12, Dec. 2018.
[55] S. J. Norton and I. J. Won, "Time exposure acoustics," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1337–1343, May 2000.
[56] J. A. Feshitan, C. C. Chen, J. J. Kwan, and M. A. Borden, "Microbubble size isolation by differential centrifugation," *J. Colloid Interface Sci.*, vol. 329, no. 2, pp. 316–324, Jan. 2009.
[57] C. C. Chen, S.-Y. Wu, J. D. Finan, B. Morrison, and E. E. Konofagou, "An experimental study on the stiffness of size-isolated microbubbles using atomic force microscopy," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 60, no. 3, pp. 524–534, Mar. 2013.
[58] S.-Y. Wu *et al.*, "Transcranial cavitation detection in primates during blood-brain barrier opening–A performance assessment study," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 61, no. 6, pp. 966–978, Jun. 2014.
[59] S.-Y. Wu, C. S. Sanchez, G. Samiotaki, A. Buch, V. P. Ferrera, and E. E. Konofagou, "Characterizing focused-ultrasound mediated drug delivery to the heterogeneous primate brain *in vivo* with acoustic monitoring," *Sci. Rep.*, vol. 6, no. 1, p. 37094, Dec. 2016.
[60] G. Pinton, J.-F. Aubry, E. Bossy, M. Muller, M. Pernot, and M. Tanter, "Attenuation, scattering, and absorption of ultrasound in the skull bone," *Med. Phys.*, vol. 39, no. 1, pp. 299–307, Dec. 2011.
[61] R. M. Jones, M. A. O'Reilly, and K. Hynynen, "Transcranial passive acoustic mapping with hemispherical sparse arrays using CT-based skull-specific aberration corrections: A simulation study," *Phys. Med. Biol.*, vol. 58, no. 14, pp. 4981–5005, Jul. 2013.
[62] C. N. Acconcia, R. M. Jones, and K. Hynynen, "Receiver array design for sonothrombolysis treatment monitoring in deep vein thrombosis," *Phys. Med. Biol.*, vol. 63, no. 23, Nov. 2018, Art. no. 235017.
[63] M. Gyöngy and C.-C. Coussios, "Passive cavitation mapping for localization and tracking of bubble dynamics," *J. Acoust. Soc. Amer.*, vol. 128, no. 4, pp. EL175–EL180, Oct. 2010.
[64] C. N. Acconcia, R. M. Jones, D. E. Goertz, M. A. O'Reilly, and K. Hynynen, "Megahertz rate, volumetric imaging of bubble clouds in sonothrombolysis using a sparse hemispherical receiver array," *Phys. Med. Biol.*, vol. 62, no. 18, pp. L31–L40, Sep. 2017.
[65] A. N. Pouliopoulos, S. Bonaccorsi, and J. J. Choi, "Exploiting flow to control the *in vitro* spatiotemporal distribution of microbubble-seeded acoustic cavitation activity in ultrasound therapy," *Phys. Med. Biol.*, vol. 59, no. 22, pp. 6941–6957, Nov. 2014.
[66] A. N. Pouliopoulos, C. Li, M. Tinguely, V. Garbin, M.-X. Tang, and J. J. Choi, "Rapid short-pulse sequences enhance the spatiotemporal uniformity of acoustically driven microbubble activity during flow conditions," *J. Acoust. Soc. Amer.*, vol. 140, no. 4, pp. 2469–2480, Oct. 2016.
[67] P. Kim, S. Bae, J. H. Song, and T.-K. Song, "Comparison study of passive acoustic mapping and high-speed photography for monitoring *in situ* cavitation bubbles," *J. Acoust. Soc. Amer.*, vol. 145, no. 6, pp. EL604–EL610, Jun. 2019.
[68] S.-Y. Wu, Y.-S. Tung, F. Marquet, C. C. Chen, and E. E. Konofagou, "Non-human primate skull effects on the cavitation detection threshold of FUS-induced blood-brain barrier opening," in *Proc. AIP Conf.*, vol. 1503, 2012, pp. 23–28.
[69] R. M. Jones and K. Hynynen, "Comparison of analytical and numerical approaches for CT-based aberration correction in transcranial passive acoustic imaging," *Phys. Med. Biol.*, vol. 61, no. 1, pp. 23–36, Jan. 2016.
[70] J. J. Choi, R. C. Carlisle, C. Coviello, L. Seymour, and C.-C. Coussios, "Non-invasive and real-time passive acoustic mapping of ultrasound-mediated drug delivery," *Phys. Med. Biol.*, vol. 59, no. 17, pp. 4861–4877, Sep. 2014.
[71] S. V. Morse *et al.*, "Rapid short-pulse ultrasound delivers drugs uniformly across the murine blood-brain barrier with negligible disruption," *Radiology*, vol. 291, no. 2, pp. 459–466, May 2019.
[72] M. A. O'Reilly and K. Hynynen, "A super-resolution ultrasound method for brain vascular mapping," *Med. Phys.*, vol. 40, no. 11, Nov. 2013, Art. no. 110701.
[73] M. A. O'Reilly, R. M. Jones, and K. Hynynen, "Three-dimensional transcranial ultrasound imaging of microbubble clouds using a sparse hemispherical array," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1285–1294, Apr. 2014.
[74] J. Li, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.
[75] P. Stoica, Z. Wang, and J. Li, "Robust capon beamforming," *IEEE Signal Process. Lett.*, vol. 10, no. 6, pp. 172–175, Jun. 2003.
[76] H. A. S. Kamimura *et al.*, "Chirp-and random-based coded ultrasonic excitation for localized blood-brain barrier opening," *Phys. Med. Biol.*, vol. 60, no. 19, pp. 7695–7712, Oct. 2015.
[77] J. Mamou, J. A. Ketterling, and R. H. Silverman, "Chirp-coded excitation imaging with a high-frequency ultrasound annular array," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 55, no. 2, pp. 508–513, Feb. 2008.
[78] A. Patel, S. J. Schoen, and C. D. Arvanitis, "Closed-loop spatial and temporal control of cavitation activity with passive acoustic mapping," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 7, pp. 2022–2031, Jul. 2019.
[79] M. E. Karakatsani, G. Samiotaki, M. Downs, V. Ferrera, and E. Konofagou, "Targeting effects on the volume of the focused-ultrasound-induced blood-brain barrier opening in non-human primates *in vivo*," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2015, pp. 3–6.

Hermes A. S. Kamimura (Member, IEEE) received the B.S. degree in medical physics and the M.S. and Ph.D. degrees in physics applied to medicine and biology from the University of São Paulo, Ribeirão Preto, Brazil, in 2008, 2011, and 2016, respectively.

He conducted research projects in therapeutic and ultrasound imaging at Mayo Clinic, Rochester, MN, USA, in 2010, and also at Columbia University, New York, NY, USA, in 2014, in student exchange programs. He was a Postdoctoral Research Scientist with French Alternative Energies and Atomic Energy Commission (CEA), Gif-sur-Yvette, France, and also with Columbia University, where he is currently an Associate Research Scientist. His research interests include harmonic motion imaging and therapeutic ultrasound spanning both ultrasound neuromodulation and ultrasound mediated blood–brain barrier disruption for targeted brain drug delivery.

Dr. Kamimura is a member of the Brazilian Physical Society and Brazilian Society of Biomedical Engineering. He was a recipient of the Outstanding Reviewer Award for Physics in Medicine and Biology, IOP Publishing in 2018, and the Best Ph.D. Dissertation Award in the field of medical physics in 2016 by the Brazilian Physical Society. He serves as a Topic Editor for Frontiers in Physics and Frontiers in Physiology.



Shih-Ying Wu (Member, IEEE) received the B.S. degree in electrical engineering from National Tsing-Hua University, Hsinchu, Taiwan, in 2007, the M.Sc. degree in biomedical electronics and bioinformatics from National Taiwan University, Taipei, Taiwan, in 2009, and the Ph.D. degree (Hons.) in biomedical engineering from Columbia University, New York, NY, USA, in 2017.

She was a Graduate Research Assistant with the Institute of Information Science, Academia Sinica, Taipei, from 2009 to 2011, before heading to the U.S. for her doctoral study. As preparation for a clinical study, she was a Postdoctoral Researcher and Translational Fellow with Columbia University, Taipei, from 2017 to 2018, developing from preclinical to clinical ultrasound technologies for blood–brain barrier opening and drug delivery. She is currently a Data Scientist with Research Computing Group, Information Services Department, Boston Children's Hospital, Boston, MA, USA.

Dr. Wu was a recipient of the Yuen-Huo Hung and Chao-Chin Huang Award of the year 2017.

Julien Grondin (Member, IEEE) received the Ph.D. degree in acoustical physics from Paris VI University (Université Pierre et Marie Curie), Paris, France, in 2010.

He is currently an Assistant Professor of radiology with Columbia University, New York, NY, USA. His research focused on bone biomechanical characterization with quantitative ultrasound. He is currently working on cardiac ultrasound imaging for myocardial elastography and electromechanical wave imaging.

Robin Ji (Graduate Student Member, IEEE) received the B.S. degree in biomedical and electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 2015. Afterwards, he joined Dr. Elisa Konofagou's Ultrasound Elasticity Imaging Laboratory at Columbia University in New York, NY, USA, where he is currently pursuing the Ph.D. degree in biomedical engineering.

His current research interest includes focused ultrasound induced blood brain barrier opening, microbubble cavitation imaging, and the development of a theranostic transducer for imaging and therapy.

Christian Aurup (Graduate Student Member, IEEE) was born in Roskilde, Denmark, in 1991. He received the bachelor's degree in biomedical engineering from the University of Delaware, Newark, DE, USA, in 2014, and the M.S. degree in biomedical engineering from Columbia University, New York, NY, USA, in 2015, where he is currently pursuing the Ph.D. degree with Ultrasound Elasticity and Imaging Laboratory (UEIL), with a focus on ultrasound-based techniques for neuromodulation and functional neuroimaging.

During his undergraduate career, he assisted with research on the DarkSide dark matter project at Princeton University, Princeton, NJ, USA, helping design water distillation techniques for removing polonium-210 from ground water and implementing them at the Laboratorio Nazionali del Gran Sasso, L'Aquila, Italy. He was an Intern with Ultrasound Elasticity and Imaging Laboratory (UEIL), Columbia University, where he worked on ultrasonic neuromodulation. His major design project was on creating a magnetic resonance (MR) compatible device for the use of lumbar elastography.

Wenlan Zheng received the B.S. degree in biomedical engineering from Columbia University, New York, NY, USA, in 2016, and the M.S. degree in public health from New York University, New York, NY, in 2019.

From 2014 to 2016, she was a Research Assistant with Ultrasound Elasticity Imaging Laboratory, Columbia University, with research interest in ultrasound-induced blood–brain barrier opening and targeted microbubble drug delivery.

Marc Heidmann received the B.S. degree in neurosciences and physics and the M.S. degree in neurosciences and cognitive engineering from École Normale Supérieure, Paris, France, in 2012 and 2014, respectively.

From 2013 to 2014, he was an Intern with Columbia University, New York, NY, USA. His research interest includes the impact of meditation on the brain plasticity, development of powerful technology to observe the brain activity and nonpharmaceutical treatment techniques against neurodegenerative diseases.

Antonios N. Pouliopoulos (Member, IEEE) was born in Thessaloniki, Greece, in 1990. He received the B.Sc. degree in physics from the Aristotle University of Thessaloniki, Thessaloniki, in 2011, the M.Sc. degree in nanotechnology and regenerative medicine from University College London, London, U.K., in 2013, and the Ph.D. degree in bioengineering from Imperial College London, London, in 2017.

As a B.Sc. student, he conducted research at the University of Bologna, Bologna, Italy, and the European Synchrotron Radiation Facility, Grenoble, France. He is currently an Associate Research Scientist with the Department of Biomedical Engineering, Columbia University, New York, NY, USA. His research interests include targeted drug delivery using ultrasound, microbubble dynamics in ultrasound therapy, and ultrasound therapy monitoring.

Elisa E. Konofagou (Senior Member, IEEE) is currently the Robert and Margaret Hariri Professor of biomedical engineering and a Professor of radiology and also the Director of the Ultrasound and Elasticity Imaging Laboratory, Biomedical Engineering Department, Columbia University, New York, NY, USA.

She has coauthored over 170 peer-reviewed journal articles. Her current research interests include the development of novel elasticity imaging techniques and therapeutic ultrasound methods and more notably, myocardial elastography, electromechanical and pulse wave imaging, harmonic motion imaging, focused ultrasound therapy, and drug delivery in the brain, with several clinical collaborations at the Columbia Presbyterian Medical Center, New York, NY, and elsewhere.

Dr. Konofagou is a Technical Committee Member of the Acoustical Society of America, the International Society of Therapeutic Ultrasound, the IEEE Engineering in Medicine and Biology Conference, the IEEE International Ultrasonics Symposium, and the American Association of Physicists in Medicine. She received the CAREER Award from the National Science Foundation and the Nagy Award from the National Institutes of Health as well as others from the American Heart Association, the Acoustical Society of America, the American Institute of Ultrasound in Medicine, the Wallace H. Coulter Foundation, the Bodossaki Foundation, the Society of Photo-Optical Instrumentation Engineers, and the Radiological Society of North America. She serves as an Associate Editor for the IEEE TRANSACTIONS ON ULTRASONICS, FERROELECTRICS, AND FREQUENCY CONTROL, ULTRASONIC IMAGING, AND MEDICAL PHYSICS.