



# Use of the cross-spectral density matrix for enhanced passive ultrasound imaging of cavitation

Maxime Polichetti, François Varray, Bruno Gilles, Jean-Christophe Béra, Barbara Nicolas

## ► To cite this version:

Maxime Polichetti, François Varray, Bruno Gilles, Jean-Christophe Béra, Barbara Nicolas. Use of the cross-spectral density matrix for enhanced passive ultrasound imaging of cavitation. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, 2020, pp.1 - 1. <10.1109/tuffc.2020.3032345>. <hal-03052463>

**HAL Id: hal-03052463**

**<https://hal.science/hal-03052463v1>**

Submitted on 10 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Use of the cross-spectral density matrix for enhanced passive ultrasound imaging of cavitation

Maxime Polichetti, François Varray, *Member, IEEE*, Bruno Gilles, Jean-Christophe Béra, and Barbara Nicolas, *Member, IEEE*,

**Abstract**—Passive ultrasound imaging is of great interest for cavitation monitoring. Spatio-temporal monitoring of cavitation bubbles in therapeutic applications is possible using an ultrasound imaging probe to passively receive the acoustic signals from the bubbles. Fourier-domain beamformers have been proposed to process the signals received into maps of the spatial localization of cavitation activity, with reduced computing times with respect to the time-domain approach, and to take advantage of frequency selectivity for cavitation regime characterization. The approaches proposed have been mainly nonadaptive, and these have suffered from low resolution and contrast, due to the many reconstruction artifacts. Inspired by the array-processing literature, and in the context of passive ultrasound imaging of cavitation, we propose here a robust estimation of the second-order statistics of data through spatial covariance matrices in the Fourier domain, or cross-spectral density matrices. The benefits of such formalism are illustrated using advanced reconstruction algorithms, such as the robust Capon beamformer, the Pisarenko class beamformer, and the multiple signal classification approach. Through both simulations and experiments in a water tank, we demonstrate that enhanced localization of cavitation activity (*i.e.*, improved resolution and contrast with respect to nonadaptive approaches) is compatible with the rapid and frequency-selective approaches of the Fourier domain. Robust estimation of the cross-spectral density matrix and the derived adaptive beamformers pave the way to the development of powerful passive ultrasound imaging tools.

**Index Terms**—Array processing, adaptive beamforming, passive ultrasound imaging, cavitation imaging

## I. INTRODUCTION

**T**herapeutic ultrasound is partly based on a complex physical phenomenon known as acoustic cavitation [1], [2]. Cavitation is known to significantly contribute in major clinical applications of high-intensity focused ultrasound (HIFU), favoring healing in ablathermy treatment of cancer [3], [4] or mechanical lysis in lithotripsy [5]. Furthermore, several prospective modalities of therapy including sonoporation and blood brain barrier opening [6], [7], sonothrombolysis [8], [9], or mechanical ablation of tissue using histotripsy [10] are based on cavitation generated by either long-pulse or short-pulse HIFU (the technique being named histotripsy in this

last case). Cavitation corresponds to the formation, oscillation, and implosion of gas bubbles in a liquid that is subjected to pressure variations induced by HIFU. As functions of the excitation conditions, different effects are induced [2]: the oscillations of the bubbles can cause tissue heating and moderate mechanical effects (*i.e.*, stable cavitation), or the implosions of the bubbles can induce destructive mechanical damage (*i.e.*, inertial cavitation). The dynamics of these bubbles are complex, nonlinear, and hard to predict, especially due to the large number of tissue characteristics involved, such that cavitation effects might appear outside of the geometric focus of the instrument during the therapy [11]. Thus, the spatio-temporal monitoring of cavitation bubbles is essential to control the actual treated area [2].

To localize cavitation, an ultrasound imaging probe composed of piezoelectric elements is used that passively receives acoustic signals from cavitation bubbles. These signals are processed via array-processing algorithms, which are also called beamformers, and which provide spatial mapping of the cavitation activity. The first method for passive cavitation imaging was proposed in [12], which was inspired by the seismic imaging method known as 'time exposure acoustics' [13], [14]. At each point of the image, it focuses, combines, and integrates the signals received over time (thus here this is referred to as TD-DAS, for 'delay-and-sum in the time-domain'). In this way, the acoustic power emitted by the medium is mapped. However, TD-DAS has three major limitations that have been frequently reported in the literature for many applications: (1) it has very low resolution, and will not always localize a small spot of cavitation precisely; (2) it does not directly differentiate stable cavitation from inertial cavitation; and (3) the focusing of the signals in the time domain is very time consuming in terms of the calculations. As a result, several variants of TD-DAS have been proposed.

On the one hand, to improve spatial localization of cavitation, adaptive methods have been proposed [15]–[17], which were inspired by active ultrasound imaging [18]–[20]. In the time domain, robust Capon beamforming (TD-RCap) [21], [22] has been adapted for passive cavitation imaging [15], [23], [24]. Based on pixel-wise optimization for combination of the delayed signals, this provides better resolution and better side-lobe suppression than the conventional approach, while being robust to model errors. However, this approach still operates in the time domain, where its optimizations makes it more time consuming than standard TD-DAS.

On the other hand, variants of TD-DAS have been proposed in the Fourier domain (FD) [25], [26]. The focusing of

M. Polichetti, F. Varray and B. Nicolas are with the University of Lyon, INSA-Lyon, UCBL, UJM-Saint-Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69621 Lyon, France.

M. Polichetti, B. Gilles and J.-C. Béra are with Inserm, U1032, LabTau, Lyon, F-69003, France; Université de Lyon, Lyon, F-69003, France; Université Lyon 1, Lyon, F-69003, France.

This study was supported by the LABEX CELYA (ANR-10-LABX-0060) and was performed within the framework of the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

signals is a simple phase shift, which significantly reduces the calculation time [27]. In addition, this facilitates frequency analysis, to distinguish between different types of cavitation. Indeed, oscillation of the stable cavitation bubbles generates narrow-band harmonic acoustic signals, while implosion of inertial cavitation bubbles creates broadband noise [28]. These approaches do not require the challenging design of pre-processing frequency filter [29], [30]. However, very few studies have been conducted to develop an adaptive beamformer in the FD for cavitation imaging.

Acoustic imaging in the FD has been widely investigated in the array-processing literature [31], [32]. Theoretical developments are often illustrated with the well-known problem of the direction-of-arrival (DOA) estimation [22], [33], [34], and have been widely applied in aeroacoustics [35]. Such studies usually represent data as a cross-spectral density matrix (CSM), which is the spatial covariance matrix of the received data at the probe expressed in the FD [31], [32], [35]. When appropriately estimated, the CSM allows access to the second-order statistics of the data, to design optimal beamformers in terms of power (e.g., minimum variance robust adaptive beamforming for improved localization performance even in presence of error models [36]), and also to use further analysis to extract source information, such as in eigenvalue decomposition of the CSM (e.g., the so-called "high-resolution" beamformers like the multiple signal classification (Music) beamformer [34]).

The goal of this paper is not to establish a definitive ranking of beamforming methods, but rather to (1) demonstrate the compatibility of adaptive beamformers and the FD; and (2) provide useful concepts of array processing and their advantages within the passive cavitation imaging field. To do so, the paper introduces a robust estimation of the CSM in the context of passive ultrasound imaging of cavitation, and in this way, proposes adaptive beamformers in the FD. Four beamformers are illustrated through the CSM formalism. First, we illustrate how the beamformer design can be derived from the CSM by implementation of the nonadaptive FD-DAS approach as a reference, as this is close to the first FD nonadaptive approach used in [25], [26], [37], and then RCap in the FD (FD-RCap), which has already been implemented in the time domain [15]. Then, the eigenvalue decomposition (EVD) of the CSM and its benefits are illustrated through the Pisarenko beamformer (FD-Pisa) [38]. Finally, taking advantage of the EVD, it is also possible to introduce another cavitation mapping approach that is not based on power estimation: the multiple signal classification beamformer (FD-Music) [34].

All of these methods have been commonly applied in many engineering fields, such as nondestructive testing, seismology, and underwater acoustics, and they are introduced in the context of passive ultrasound imaging. They are first illustrated in simulations. Then, under experimental conditions, we demonstrate their potential for spatio-temporal monitoring of cavitation for real cavitation bubbles.

## II. MODEL AND ESTIMATION OF THE CROSS-SPECTRAL DENSITY MATRIX

This section first introduces the signal models in both the time and frequency domains. Then the CSM and its characteristics are defined. Finally, the robust estimation of the CSM is proposed.

### A. Signal model

An acoustic source localized in  $\vec{r}_0$  emits a real signal  $s(t)$ . This is passively received by a uniform linear array of  $N$  sensors, where  $n$  is the sensor index ( $n = 1, \dots, N$ ). Assuming the source transmits spherical waves and the sound speed  $c$  is uniform in the propagation medium, the time of flight from the source to each element of the probe is

$$\tau_n(\vec{r}_0) = \frac{\|\vec{r}_n - \vec{r}_0\|}{c}. \quad (1)$$

Assuming there is no attenuation of the wave, the data matrix that contains the received signals is given by [15], [26]:

$$\mathbf{y}(t) = [s(t - \tau_1(\vec{r}_0)), \dots, s(t - \tau_N(\vec{r}_0))]^T + \mathbf{b}(t), \quad (2)$$

where  $^T$  is the transposition operator, and  $\mathbf{b}(t)$  is the measurement noise. This  $\mathbf{b}(t)$  should be white Gaussian noise, with zero-mean, and also spatially white. The data matrix  $\mathbf{y}(t)$  is of size  $N \times N_t$ , where  $N_t$  is the number of time samples (which corresponds to an acquisition duration of  $T_{rec}$ ). In the FD, for a given frequency  $f$ , the received data are represented by an  $N$ -length column vector as [26], [33]:

$$\mathbf{Y}(f) = S(f) \mathbf{a}(\vec{r}_0, f) + \mathbf{B}(f), \quad (3)$$

where  $S(f)$  is the Fourier transform of the source signal  $s(t)$ ,  $\mathbf{B}(f)$  is the vector of the Fourier transforms of the noise contributions for each sensor, and  $\mathbf{a}(\vec{r}_0, f)$  is the propagation vector, such that:

$$\mathbf{a}(\vec{r}_0, f) = [e^{-i2\pi f \tau_1(\vec{r}_0)}, \dots, e^{-i2\pi f \tau_N(\vec{r}_0)}]^T. \quad (4)$$

Acoustic imaging aims to build maps on the basis of received signals (3), on which the peak values correspond to the positions of the sources. To do this, spatial filters, called beamformers, have been designed to detect the spatial signature  $\mathbf{a}(\vec{r}_0, f)$  of the array of an acoustic source placed at  $\vec{r}_0$ . More precisely, beamforming aims to detect the relative phase shifts between channels due to the delays  $\tau_n(\vec{r}_0)$ .

### B. Cross-spectral density matrix analysis

Beamforming design is often based on the spatial covariance matrix of the data received, for three main reasons [31], [32], [39]. First, this matrix directly expresses the phase shifts from one element to another of the signals received, and thus the spatial signature of the acoustic source. Secondly, if estimated in the correct way, this gives access to the second-order statistics of the received data, which allows the optimal beamforming design (in terms of power or variance). Thirdly, the eigenstructure of the data can be used to extract the source information. In the first subsection II-B1, the CSM is defined. Then, subsections II-B2 and II-B3 introduce the notations and concepts derived from the CSM, which are required to understand the four beamformers considered.

1) *Definition:* Signals are considered broadband, zero-mean, stationary, and ergodic, to deal with their second-order statistics. In the FD, the spatial covariance matrix is usually referred to as the CSM. The CSM is of size  $N \times N$ , and it is defined as:

$$\mathbf{M}(f) = E \left\langle \lim_{T_{int} \rightarrow +\infty} \frac{1}{T_{int}} \mathbf{Y}(f) \mathbf{Y}^*(f) \right\rangle, \quad (5)$$

where  $E \langle \cdot \rangle$  is the expected value,  $T_{int}$  is the integration time for the Fourier analysis, and  $*$  is the transpose conjugate operator. The diagonal coefficients are the power auto-spectrum densities of the channels. The other coefficients are the power cross-spectrum densities, which reveal the relative phase-shifts between the channels.

2) *Use of the cross-spectral density matrix for power estimation:* Passive ultrasound imaging of cavitation was essentially developed through reconstruction of power maps [25], [26]. Generally speaking, the FD beamformers are spatial filters that consist of applying  $N$  complex coefficients, contained in the so-called steering vector  $\mathbf{h}(\vec{r}, f)$ , to the observations  $\mathbf{Y}(f)$ , to compensate for the phase shifts induced by the propagation vector  $\mathbf{a}(\vec{r}_0, f)$  on the signals received, and then coherently sum them. The mean power spectral density for the random signal  $\mathbf{h}^*(\vec{r}, f) \mathbf{Y}(f)$  received at the probe and steered at  $\vec{r}$  is expressed as:

$$P(\vec{r}, f) = E \left\langle \lim_{T_{int} \rightarrow +\infty} \frac{1}{T_{int}} |\mathbf{h}^*(\vec{r}, f) \mathbf{Y}(f)|^2 \right\rangle. \quad (6)$$

This expression is looked upon as a power estimate, as commonly used for passive cavitation imaging, although with a statistical point of view, it can be rewritten using the CSM from Equation (5):

$$P(\vec{r}, f) = \mathbf{h}^*(\vec{r}, f) \mathbf{M}(f) \mathbf{h}(\vec{r}, f). \quad (7)$$

The advantage of this formalism compared to the methods used in [25], [26] is that power estimate at a given position  $\vec{r}$  is expressed in terms of the second-order statistics of the data, which gives access to optimal beamforming in terms of power or variance.

To distinguish the source from the noise contributions, and thus to design the optimal steering vector, using Equations (3) and (5), Equation (7) can be developed as:

$$P(\vec{r}, f) = |\mathbf{h}^*(\vec{r}, f) \mathbf{a}(\vec{r}_0, f)|^2 P_s(f) + \mathbf{h}^*(\vec{r}, f) \mathbf{M}_b(f) \mathbf{h}(\vec{r}, f), \quad (8)$$

where  $P_s(f)$  is the power spectral density of the source, and  $\mathbf{M}_b(f)$  is the CSM of the noise. Here, the quantity  $G_{\vec{r}}(\vec{r}_0, f) = |\mathbf{h}^*(\vec{r}, f) \mathbf{a}(\vec{r}_0, f)|^2$  is the receive spatial gain of the beamforming method. The benefits of this formalism are illustrated later for FD-DAS (III-A) and FD-RCap (III-B).

3) *Eigenvalue decomposition of the cross-spectral density matrix:* The CSM gives access to very interesting representations

of the received data through the EVD of the matrix  $\mathbf{M}(f)$ , which is defined as [31], [35]:

$$\begin{aligned} \mathbf{M}(f) &= \mathbf{U}(f) \mathbf{V}(f) \mathbf{U}^*(f) \\ &= \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_N \end{bmatrix} \begin{bmatrix} v_1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & v_N \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^* \\ \vdots \\ \mathbf{u}_N^* \end{bmatrix} \\ &= \sum_{n=1}^N v_n \mathbf{u}_n \mathbf{u}_n^*, \end{aligned} \quad (9)$$

where  $\mathbf{V}(f)$  and  $\mathbf{U}(f)$  are the matrices that contain the eigenvalues (EVs)  $v_n$  and the column eigenvectors  $\mathbf{u}_n$ , respectively. The EVs are sorted in decreasing order ( $v_1 > \dots > v_N$ ). In a noiseless case,  $J$  sources are said to be spatially decorrelated if they are orthogonal in the sensor space, i.e., the (normalized) propagation vector of the source  $j$  corresponds to a CSM eigenvector:  $\mathbf{u}_j = \mathbf{a}(\vec{r}_j, f) / \sqrt{N}$ . In this case, each EV  $v_j$  is  $N$  times the power spectral density  $P_{s,j}(f)$  of the source  $j$ , as a signal received on the array is recorded  $N$  times:  $v_j = N P_{s,j}(f)$ .

In the presence of spatially decorrelated background noise, e.g.,  $\mathbf{M}_b(f) = \sigma_b^2 \mathbf{I}$ , where  $\sigma_b^2(f)$  is its power spectral density and  $\mathbf{I}$  is the identity matrix, the decomposition is modified. The EVD cannot perfectly separate the sources  $j$  from the noise for finite-size observations (e.g., number of sensors  $N$  is finite). This is uniformly distributed over all EVs. The CSM is decomposed as the sum of a signal space  $\mathbf{E}_s$  that is composed of the first  $J$  eigenvectors, and a noise space  $\mathbf{E}_b$  that is composed of the remaining eigenvectors, such that:

$$\begin{aligned} \mathbf{M}(f) &= \underbrace{\mathbf{E}_s}_{\sum_{n=1}^J (N P_{s,n}(f) + \sigma_b^2) \mathbf{u}_n \mathbf{u}_n^*} + \underbrace{\mathbf{E}_b}_{\sum_{n=J+1}^N \sigma_b^2 \mathbf{u}_n \mathbf{u}_n^*} \\ &= \sum_{n=1}^J (N P_{s,n}(f) + \sigma_b^2) \mathbf{u}_n \mathbf{u}_n^* + \sum_{n=J+1}^N \sigma_b^2 \mathbf{u}_n \mathbf{u}_n^*. \end{aligned} \quad (10)$$

Each source  $j$  is decomposed on one of the subspaces  $\mathbf{u}_n \mathbf{u}_n^*$  that forms the signal space  $\mathbf{E}_s$ , and is orthogonal to the other subspaces, and notably those of the noise space  $\mathbf{E}_b$ . The EVD gives access to new beamformers for passive ultrasound imaging, as illustrated in III-C and III-D.

### C. Cross-spectral density matrix estimation

The quality of the cavitation maps greatly depends on how the matrix  $\mathbf{M}(f)$  is estimated, and the way that it is further processed. Assuming the received signals are ergodic and stationary, a natural way to estimate  $\mathbf{M}(f)$  is to consider  $K$  successive acquisitions  $\mathbf{y}_k(t)$  (usually called snapshots) of duration  $T_{snap}$ :

$$\hat{\mathbf{M}}(f) = \frac{1}{K} \sum_{k=1}^K \frac{1}{T_{snap}} \mathbf{Y}_k(f) \mathbf{Y}_k^*(f) \quad (11)$$

where  $\mathbf{Y}_k(f)$  is the Fourier transform of the  $k^{th}$  snapshots  $\mathbf{y}_k(t)$ . Beamformer performances increase with  $K$  [22]. However,  $T_{snap}$  has to be high enough to ensure good spectral resolution. To artificially increase the number of realizations  $K$  without extensively reducing  $T_{snap}$ , a single acquisition

TABLE I: Summary of the beamformer characteristics.

Name	Adaptive parameter	Power-estimate based	Computing cost
FD-DAS (Delay-And-Sum)	--	Yes	Low
FD-Rcap (Robust Capon)	$\epsilon$ (tolerance to uncertainties)	Yes	High
FD-Pisa (Pisarenko class)	$p$ ( $p^{\text{th}}$ root compression of eigenvalues)	Yes	Low
FD-Music (Multiple Signal Classification)	$K_s$ (estimated signal space length)	No	Low

of length  $T_{rec}$  is divided into partially overlapping snapshots  $\mathbf{y}_k(t)$ . The overlap  $p_{ov}$  is expressed as a percentage of  $T_{snap}$ . The parameters to estimate the CSM are then linked by the following relationship:

$$K = \frac{T_{rec}}{(1 - p_{ov})T_{snap}}. \quad (12)$$

### III. BEAMFORMING METHODS

A wide variety of beamformers has been proposed in the array-processing literature [31], [32], [35]. To illustrate the advantages of the proposed formalism, four beamformers that use the second-order statistics of the received data through the CSM are introduced within the framework of passive ultrasound imaging. First, we extend two already existing beamformers for passive cavitation imaging into the proposed formalism: FD-DAS is an extension of [25], but includes the robust estimate of CSM; and FD-RCap corresponds to the beamformer of [15], but transposed into the FD. In this way, we present nonadaptive and adaptive ways, respectively, to use the CSM for optimal steering-vector design. Then, the proposed formalism is used to introduce new beamformers for passive cavitation imaging. In particular, the benefits of exploitation of the eigenstructure of the received data are illustrated for two advanced beamformers. The FD-Pisa is an intermediate approach between FD-DAS and FD-RCap in terms of map quality and complexity. Unlike the other three beamformers, FD-Music is not a power-estimate-based method, although it is well known in the field of array processing for its high-resolution performance. A summary of the beamformer characteristics is given in Table I. Note that although the spectral signature of cavitation is broad, the following analyses consider single-frequency imaging, as is often the case in the literature [26], [27]. The perspectives for composite multifrequency imaging are further discussed in section VI-4.

#### A. The nonadaptive beamformer: FD-DAS

The formulation of FD-DAS can be described as a minimization of a cost function, as for DOA estimation in [40]. The design of the steering vector  $\mathbf{h}(\vec{r}, f)$  is here independent of the signals received. First, its coefficients have to minimize the contribution of the noise, and in the basic case it is spatially uncorrelated, such that  $\mathbf{M}_b(f) = \sigma_b^2 \mathbf{I}$ . Secondly, for each pixel in  $\vec{r}$ , the coefficients in  $\mathbf{h}(\vec{r}, f)$  must preserve the signal that

would have been emitted with the spatial signature  $\mathbf{a}(\vec{r}, f)$ , with a unit gain  $G_{\vec{r}}(\vec{r}, f) = 1$ . From Equation (8), the problem can be written as:

$$\begin{aligned} \min_{\mathbf{h}(\vec{r}, f)} \quad & \mathbf{h}^*(\vec{r}, f) \mathbf{h}(\vec{r}, f) \\ \text{subject to} \quad & \mathbf{h}^*(\vec{r}, f) \mathbf{a}(\vec{r}, f) = 1. \end{aligned} \quad (13)$$

The analytical solution is obtained by the Lagrange multipliers method:

$$\mathbf{h}_{FD-DAS}(\vec{r}, f) = \frac{\mathbf{a}(\vec{r}, f)}{N}. \quad (14)$$

By introducing Equation (14) into the power spectral density estimation of Equation (7), and replacing  $\mathbf{M}(f)$  by its estimate  $\hat{\mathbf{M}}(f)$ , this gives:

$$\hat{P}_{FD-DAS}(\vec{r}, f) = \frac{\mathbf{a}^*(\vec{r}, f) \hat{\mathbf{M}}(f) \mathbf{a}(\vec{r}, f)}{N^2}. \quad (15)$$

This expression provides access to a nonadaptive power-estimate-based beamformer with a robust estimation of the CSM using snapshot averaging. Note that when  $K = 1$  (i.e., without block averaging for the CSM estimation; see Equation (11)), this corresponds to [25], [26]. The nonadaptive approach is very simple and quick, but it is well-known for its poor resolution and for numerous artifacts [25], [26].

#### B. The Capon adaptive beamformers: FD-Cap, FD-RCap

Capon adaptive beamforming [40], [41] aims to optimize the design of the steering vector for each pixel using the information contained in the received data, via the matrix  $\mathbf{M}(f)$ . In this way, acoustic sources are better resolved and artifact rejection is improved, with respect to the nonadaptive approach.

With a unit gain constraint for the signal that would have been emitted from  $\vec{r}$  (i.e., the same constraint as for FD-DAS), FD-Cap proposes to find the optimal steering vector that minimizes the power of all types of noise (i.e., without assumptions as to the structure of its CSM  $\mathbf{M}_b(f)$ ; unlike FD-DAS). Although it is not possible to estimate this directly, in practice  $\mathbf{M}_b(f)$  is proportional to  $\mathbf{M}(f)$  [36]. The problem of minimization of FD-Cap is then written as:

$$\begin{aligned} \min_{\mathbf{h}(\vec{r}, f)} \quad & \mathbf{h}^*(\vec{r}, f) \mathbf{M}(f) \mathbf{h}(\vec{r}, f) \\ \text{subject to} \quad & \mathbf{h}^*(\vec{r}, f) \mathbf{a}(\vec{r}, f) = 1. \end{aligned} \quad (16)$$

This problem has an analytical solution, which can be solved using the Lagrange multipliers method:

$$\mathbf{h}_{FD-Cap}(\vec{r}, f) = \frac{\mathbf{M}^{-1}(f) \mathbf{a}(\vec{r}, f)}{\mathbf{a}^*(\vec{r}, f) \mathbf{M}^{-1}(f) \mathbf{a}(\vec{r}, f)}. \quad (17)$$

Introduction of Equation (17) into Equation (7), and replacing  $\mathbf{M}(f)$  by its estimate  $\hat{\mathbf{M}}(f)$ , gives:

$$\hat{P}_{FD-Cap}(\vec{r}, f) = \frac{1}{\mathbf{a}^*(\vec{r}, f) \hat{\mathbf{M}}^{-1}(f) \mathbf{a}(\vec{r}, f)}. \quad (18)$$

The Capon approach is not used in practice, because it is very unstable in the presence of model errors [15]. Many approaches for telecommunications or acoustic imaging have been developed to make this method more robust [18], [36].

The FD-RCap [21], [22], [33] considers that if there are modeling errors (*i.e.*, sound speed, probe calibration), then the true propagation vector  $\mathbf{a}_0(\vec{r}, f)$  is included in a small region around the  $\mathbf{a}(\vec{r}, f)$  vector, which is defined by a parameter  $\epsilon$ . The FD-RCap looks for the optimal estimation  $\hat{\mathbf{a}}_0(\vec{r}, f)$  of  $\mathbf{a}_0(\vec{r}, f)$  from the received data. The optimization problem arises as

$$\begin{aligned} \min_{\hat{\mathbf{a}}_0(\vec{r}, f)} \quad & \hat{\mathbf{a}}_0^*(\vec{r}, f) \mathbf{M}^{-1}(f) \hat{\mathbf{a}}_0(\vec{r}, f) \\ \text{subject to} \quad & \|\hat{\mathbf{a}}_0(\vec{r}, f) - \mathbf{a}(\vec{r}, f)\|^2 \leq \epsilon. \end{aligned} \quad (19)$$

The top term here tries to maximize the power estimation in the sense of Capon (Equation (18)) and the bottom term defines the uncertainty region around  $\mathbf{a}(\vec{r}, f)$ . This problem has an analytical solution that can also be obtained by the Lagrange multipliers method:

$$\hat{\mathbf{a}}_0(\vec{r}, f) = \mathbf{a}(\vec{r}, f) - (\mathbf{I} + \lambda(\epsilon) \hat{\mathbf{M}}(f))^{-1} \mathbf{a}(\vec{r}, f), \quad (20)$$

where  $\mathbf{I}$  is the identity matrix, and  $\lambda(\epsilon)$  is the Lagrange multiplier that is optimally computed for each pixel using a Newton-Raphson numerical method that is controlled by the  $\epsilon$  parameter. Introducing Equation (20) into the Capon solution of Equation (18), and simplifying [22], this gives

$$\hat{P}_{FD-RCap}(\vec{r}, f) = \frac{\|\hat{\mathbf{a}}_0(\vec{r}, f)\|^2/N}{\hat{\mathbf{a}}_0^*(\vec{r}, f) \hat{\mathbf{M}}^{-1}(f) \hat{\mathbf{a}}_0(\vec{r}, f)}. \quad (21)$$

The  $\epsilon$  parameter depends on the received data and regulates the compromise between the robustness and the image quality. A too low value of  $\epsilon$  defines a too small region of uncertainty, whereby FD-RCap tends towards FD-Cap, and thus loses robustness: at the slightest model error, the source signal is interpreted as noise and is rejected. A too high value of  $\epsilon$  defines a too large region of uncertainty: the power of the source is overestimated, and the interference is amplified. The choice of this parameter is generally empirical, with the condition  $\epsilon < \|\mathbf{a}(\vec{r}, f)\|^2$ , and it is not a big issue as it can be found by a little trial and error [15], [21]. It should be noted that the inversion of  $\hat{\mathbf{M}}(f)$  implies the matrix is well conditioned, with a sufficient condition of  $K > N$  [22], [40].

### C. The Pisarenko class beamformer: FD-Pisa

An alternative to the pixel-wise optimal design of the steering vector is to artificially change the relative levels between the sources prior to steering. This is possible by taking advantage of the eigenstructure of  $\mathbf{M}(f)$  and applying nonlinear operations to its EVs. This approach is known as Pisarenko class beamforming [38], [42], [43], or FD-Pisa. Namely, to sharpen the peaks that reflect the source positions (with respect to FD-DAS), the  $p^{th}$  root is applied to the EVs of  $\mathbf{M}(f)$ :

$$\mathbf{M}^{1/p}(f) = \sum_{n=1}^N v_n^{1/p} \mathbf{u}_n \mathbf{u}_n^*. \quad (22)$$

In this way, the gap between the source levels is compressed: the sidelobes generated by strong sources will interfere less with the lower source level when beamforming. The steering

vector is the same as for FD-DAS, with a previous normalization such that no bias is introduced when using nonlinear operations on it,

$$\mathbf{h}_{FD-Pisa}(\vec{r}, f) = \frac{\mathbf{a}(\vec{r}, f)}{\sqrt{N}}. \quad (23)$$

After steering the compressed CSM  $\mathbf{M}^{1/p}(f)$ , the  $p$  power is used on the resulting value to reestablish the original dimension of an acoustic power:

$$\hat{P}_{FD-Pisa}(\vec{r}, f) = \frac{1}{N} \left( \frac{\mathbf{a}^*(\vec{r}, f)}{\sqrt{N}} \hat{\mathbf{M}}^{1/p}(f) \frac{\mathbf{a}(\vec{r}, f)}{\sqrt{N}} \right)^p \quad (24)$$

where the  $1/N$  term scales the  $N$  factor within the EVD (see Equation (10) and [38]). The parameter  $p$  has to be empirically chosen:  $p = 1$  corresponds to FD-DAS, and  $p > 1$  is interesting to sharpen the peak values. The performances in terms of localization usually increase with  $p$  when uncorrelated sources are considered [38], [43], [44].

### D. The multiple signal classification beamformer: FD-Music

The FD-Music [34], [45] is not power-estimation based, but it is well known as a high resolution beamformer. FD-Music localizes the sources using the contributions of the background noise, rather than the source signal directly. More precisely, FD-Music exploits the orthogonality between the signal subspace and the noise subspace obtained by EVD decomposition of  $\mathbf{M}(f)$ , as described in Equation (10).

For each pixel  $\vec{r}$  of the map, FD-Music projects the steering vector  $\mathbf{h}_{FD-Music}(\vec{r}, f) = \mathbf{a}(\vec{r}_j, f)$  to the noise subspace: this is the sum of the scalar products of  $\mathbf{h}_{FD-Music}(\vec{r}, f)$  with each vector of  $\mathbf{E}_b$ . The steering vector is chosen proportional to that of the FD-DAS in Equation (14), without the  $1/N$  factor, as the FD-Music does not seek to quantify the power. Two cases are then possible:

- The pixel in  $\vec{r}$  coincides with the position  $\vec{r}_j$  of the source  $j$ . Then the pointing vector  $\mathbf{h}_{FD-Music}(\vec{r}, f) = \mathbf{a}(\vec{r}_j, f)$  is a eigenvector of the signal space  $\mathbf{E}_s$ . By construction of the EVD, it is therefore necessarily orthogonal to each eigenvector of the noise space  $\mathbf{E}_b$ , and therefore the sum of these scalar products is also zero:

$$\sum_{n=J+1}^N \mathbf{a}^*(\vec{r}_j, f) \cdot \mathbf{u}_n = 0. \quad (25)$$

- The pixel in  $\vec{r} \neq \vec{r}_j$  does not coincide with any of the  $J$  sources. So the steering vector  $\mathbf{h}_{FD-Music}(\vec{r}, f) = \mathbf{a}(\vec{r}, f)$  is not an eigenvector of  $\mathbf{E}_s$ . There is, therefore, at least one eigenvector of the noise space for which its projection is not zero. So by necessity, one of the scalar products  $\mathbf{a}^*(\vec{r}, f) \cdot \mathbf{u}_n$  ( $\forall n \in \llbracket J + 1 ; N \rrbracket$ ) is not zero. Thus the sum of the scalar products is not zero:

$$\sum_{n=J+1}^N \mathbf{a}^*(\vec{r}, f) \cdot \mathbf{u}_n \neq 0, \quad \text{if } \vec{r} \neq \vec{r}_j. \quad (26)$$

Taking advantage of the properties defined by Equations (25) and (26), the FD-Music map is defined by:

$$\hat{P}_{FD-Music}(\vec{r}, f) = \frac{1}{\sum_{n=K_s+1}^N |\mathbf{a}^*(\vec{r}, f) \cdot \mathbf{u}_n|^2}, \quad (27)$$

where  $K_s$  is an estimation of the signal space size, as the number of sources  $J$  is unknown in practice. The choice of size  $K_s$  for  $\mathbf{E}_s$  is often guided by analysis of the decrease in EVs [46], [47]. The maxima of  $\hat{P}_{FD-Music}(\vec{r}, f)$  are reached when its denominator tends toward zero (equivalent to Equation (25)), and therefore indicates the source position. Indeed, in practice, on a source pixel, the denominator does not cancel itself out strictly, but tends toward 0, giving peak values on the map. For the sake of uniformity in localization map notation, it can be noted  $\hat{P}_{FD-Music}(\vec{r}, f)$ , although it has to be kept in mind that it is not a power map.

#### IV. MATERIALS

To illustrate the relevance of this exploitation of the robust estimation of the CSM and the resulting FD adaptive beamformers within the passive cavitation imaging framework, single and multiple source configurations were carried out on simulations, and an experimental set-up was developed to create cavitation bubbles in a water tank. We first describe this set-up, and then the simulations using similar parameters. The performance metrics are also detailed.

##### A. Experimental set-up

1) *High-intensity focused ultrasound settings*: The experimental set-up is represented in Fig. 1. Cavitation bubble clouds were created using a 1-MHz single spherical transducer focused at 8 cm, with an active diameter of 10 cm, and a circular 4.5-cm diameter aperture in the center (Imasonic, France). This was immersed in a tank of filtered water (72 cm×25 cm×25 cm). The theoretical focal volume at -3 dB is ellipsoidal in shape, 8.3-mm long for the  $x$ -axis, and 1.1-mm in diameter. The HIFU transducer is connected to a power amplifier (GN1000; PRANA, France). This can transmit bursts centered at 1 MHz, with each lasting 20 ms, and pulsed at a rate of 4 Hz. The amplitude delivered is adjustable, although it remains constant during each shot. The amplitudes can range from 10 mV to 150 mV, which generates peak negative pressures at the focus of the transducer of 0.72 MPa to 10.8 MPa. Here, we focus on the data at 7.2 MPa.

2) *Passive ultrasound imaging*: For passive ultrasound imaging, a linear ultrasound imaging probe (L7-4; Verasonics, USA) was used. It is composed of  $N = 128$  elements with uniform spacing of 298  $\mu\text{m}$ , for a total width of 38 mm. Its center frequency is 5 MHz and its bandwidth at -6 dB is 3.5 - 7.5 MHz. It is connected to an ultrasound scanner (Vantage 256; Verasonics, USA). The probe makes continuous passive acquisitions with a duration of 24 ms at a sampling frequency of  $f_s = 20.8$  MHz. Although the therapy signal oscillates at a fundamental frequency of  $f_{HIFU} = 1$  MHz, outside the bandwidth of the imaging probe the images are generally reconstructed by exploiting higher frequencies than for cavitation, which are within the bandwidth of the imaging probe, to ensure better resolution [26], [30], [37], [48]. The imaging probe is placed 6 cm above the focus of the HIFU transducer. The lateral axis of the imaging probe is parallel to the  $x$ -axis, and points toward the HIFU transducer. The ultrasound images are reconstructed in the  $xz$ -plane.

3) *Optical imaging*: Ombroscopic imaging of the cavitation phenomenon was obtained using a fast optical camera (Phantom v12.1; Vision Research Inc., USA) that recorded at the rate of 10 000 frames per second, and an exposure time of 99  $\mu\text{s}$ , over 24 ms. The images obtained are in the same  $xz$ -plane as the ultrasound images. The field of view is 1.5-cm wide and 1.1-cm high (800×600 pixels), and contains the HIFU focus.

4) *Cavitation generation*: The objective was to compare optical images with the reconstructed passive ultrasound maps of cavitation events that consisted of a few bubbles nucleated with the HIFU bursts. Optical images are limited to a very small region compared to the broad area where the cavitation might be nucleated and within which this might move in the water tank (not necessarily near the focus). To ensure comparison with optical images, *i.e.*, to trap the cavitation within the field of view of the camera, a polyester wire with asperities that provided nucleation sites along the wire was placed at the HIFU focus (Fig. 1). In this way, the cavitation was easier to image optically, as it occurs mainly in the neighborhood of the intersection of the wire and the HIFU focus. Also, the cavitation threshold was stabilized, and essentially depended on the gas concentration (estimated from the oxygen concentration). The cavitation threshold was measured using spectrogram analysis of the signal passively recorded with the central element of the ultrasound probe, as was done with a hydrophone in [28], [49]. Two oxygen concentrations were used: 2.5 mg/L and 6.0 mg/L, which corresponded to the measured cavitation thresholds of 5 MPa and below 0.7 MPa, respectively. This experimental set-up allowed comparison of the optical and ultrasound imaging modalities of a wide variety of cavitation events. In this paper, we focus on the cavitation events that consisted of the few nucleated bubbles that stayed at the wire surface. In a second configuration, a second wire (wire #2) was placed 3-mm downstream from the first one (wire #1); see Fig. 1. This wire #2 was still located within the -3 dB focal volume, although at a point where the HIFU field was less intense, and thus this produced a second weaker source in comparison to the first source created on wire #1, which constitutes a more challenging scenario to be imaged, with artifacts of the strong source spreading on the weaker one. In the results section, the presence of inertial cavitation through the whole HIFU burst was confirmed using spectrogram analysis, as in [28], [49].

5) *Synchronization and registration of images*: The control system for the HIFU transducer generates a synchronization signal 1 ms before the start of the HIFU shot. This signal is received by the ultrasound scanner and launches passive acquisition for a period of 24 ms. The same applies to the optical camera. The ultrasound propagation and device activation times are less than 100  $\mu\text{s}$ . Thus, cavitation is imaged before, during, and after the entire shot.

The ultrasound and optical images are spatially registered using two nodes on the wire as markers. Before the HIFU shot, these are localized by active ultrasound imaging. Their coordinates are then assigned to the nodes visible in the optical image. In this way, the optical images are described via the  $x$  and  $z$  axes, as with the ultrasound images. Comparison is then



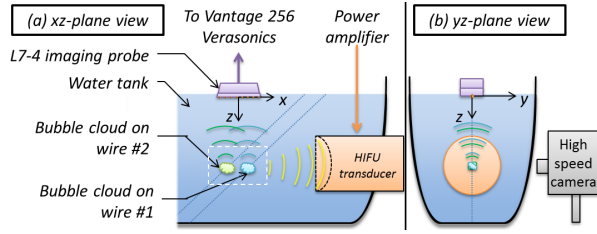


Fig. 1: Experimental set-up showing the  $xz$ -plane view (a) and the  $yz$ -plane view (b). Wire #1 is placed in the  $xz$ -plane, to cross the acoustic axis at the theoretical focus of the HIFU transducer. HIFUs propagate in the opposite direction to the  $x$ -axis. When the HIFU transducer fires, a cavitation cloud is created: the acoustic signal emitted by the bubbles is passively received by the imaging probe. In the second configuration, a second wire #2 is added 3 mm downstream of wire 1, and parallel to it. In (a), white dots indicate the portion of the  $xz$ -plane that is imaged by the high-speed camera.

possible with maps obtained by passive ultrasound imaging.

6) *Reconstruction parameters:* For passive ultrasound imaging, spatio-temporal tracking of the cavitation is ensured by dividing the total acquisition of 24 ms into 120 sub-acquisitions of  $T_{rec} = 200 \mu s$ , with each of these providing a localization map. For each map, snapshot averaging is achieved with  $K = 130$  realizations and  $T_{snap} \simeq 15 \mu s$ ; *i.e.*, with an overlap of  $p_{ov} = 90 \%$  (as the parameters that provided the best results in the simulation V-A3).

### B. Simulation set-up

Simulations are conducted with probe parameters similar to those described in section IV-A2, except for the frequency response of the probe, which is not simulated. Four simulated configurations are considered, as summarized in Figure 2. All of the configurations consider sources that emit random Gaussian noise to mimic cavitation noise, as proposed in [29], through the whole acquisition, which lasts for  $T_{rec} = 200 \mu s$ . To mimic the measurement noise, white Gaussian noise is systematically added, independently to each sensor, with a varying signal-to-noise ratio (SNR).

Configuration 1 is designed to investigate the evolution of the common image quality metrics (*i.e.*, lateral and axial main lobe widths, contrast, and error position, as detailed in section IV-C) with respect to the SNR varying between  $-15$  dB and  $+15$  dB. To do so, a single source is placed at  $\vec{r}_1 = [5, 60]$  mm.

Configurations 2 and 3 are carried out to illustrate the superiority of adaptive beamformers to resolve two closely spaced sources for the lateral and axial directions, respectively. To do so, a first source is placed at  $\vec{r}_1 = [5, 60]$  mm, and a second one is added, for configuration 2 with 0.4 mm lateral spacing (about one wavelength  $\lambda = 0.31$  mm), and for configuration 3 with 3.5 mm axial spacing (about 11 wavelengths). In both configurations, the two sources have the same transmitting amplitude, with a SNR of 10 dB.

Configuration 4 is simulated to highlight the influence of CSM averaging on cavitation maps; namely, in a complex scenario where three sources coexist. Moreover, configuration 4

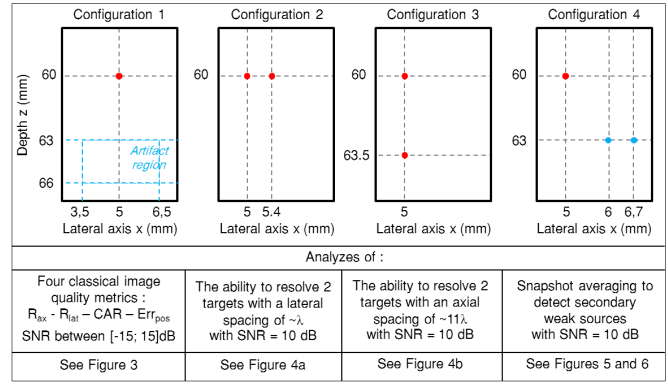


Fig. 2: Descriptions of the four simulated configurations and their respective goals. Note that the wavelength is  $\lambda = 0.31$  mm.

investigates the interest in the use of FD adaptive beamformers to reduce the side lobes of a strong source that hides two weaker sources. To do so, three sources are simulated, with the first high-intensity source at  $\vec{r}_1 = [5, 60]$  mm. The second source is placed in the side lobes generated by the FD-DAS at  $\vec{r}_2 = [6, 63]$  mm, with a power of 10 dB less than the first one. The third source is placed close to the second source, at  $\vec{r}_3 = [6.7, 63]$  mm (with an approximately two-wavelength spacing), and also with 10 dB less power than the first one. The SNR is 10 dB.

### C. Beamformer performance metrics

The benefits of the CSM formalism are assessed through qualitative and quantitative comparisons of the imaging methods. Simple metrics commonly used in the context of ultrasound imaging are considered, to characterize the resolution and contrast of the images obtained for both the simulated and experimental data.

The full-width at half-maximum of the main lobe is measured in millimeters, as  $R_{ax}$  in the axial dimension and  $R_{lat}$  in the lateral dimension. This is a simple metric measured on single-source configuration (or isolated source), as commonly used to reflect the resolution of two sources by a beamformer.

The gap measurement between two close sources represents a more relevant metric to quantify the resolution performance of a beamformer, although this is more complex. It is computed as (before log-compression of the map values):

$$Gap = 10 \log_{10} \left( \frac{\max(\hat{P}(\vec{r}_1), \hat{P}(\vec{r}_2))}{\hat{P}(\frac{\vec{r}_1 + \vec{r}_2}{2})} \right). \quad (28)$$

The cavitation-to-artifact ratio ( $CAR$ ) is defined as the ratio of the maximum value from the map over the averaged pixel values in an artifact region (before log-compression of the map values). The  $CAR$  measurement is a value with no physical meaning, but that aims to provide a relative comparison of the rejection of artifacts by each beamformer (*e.g.* side lobes, main lobe extension under  $-3$  dB). The  $CAR$  reflects the contrast, and thus the unmasking of weak sources by the beamformer in the presence of stronger sources. The position and the extent of



the artifact region do not favour a beamformer (except maybe in a very pathological case, as very far from the target, or including the main lobe above  $-3$  dB).

The position error  $Err_{pos}$  is the distance in millimeters between the actual position of a simulated point source and the maximum pixel value in the map (for a single source scenario).

The running time for the reconstruction is measured to reflect the complexities of the beamformers. All of the reconstructions are carried out in Matlab (R2016b, The MathWorks, Natick, MA, USA), on the same computer, which has 32 GB memory and an Intel(R) Core(TM) i7-6600 CPU at 2.60 GHz.

## V. RESULTS

We first present the data for the simulations, to introduce the performances of the four beamformers and the influence of the robust CSM estimation for the different point source configurations (Fig. 2). Then, the experimental data for small cavitation clouds are described. According to the guidelines given in the literature, the adaptive parameters for FD-RCap [15], [21] and FD-Pisa [43] are empirically chosen to satisfy the compromise between resolution and contrast (as indicated for each configuration), as they are not easily physically interpretable. For FD-Music, the choice of  $K_s$  is more guided (as it corresponds to the number of uncorrelated sources, which is easy to know in simulation, and also due to the EV plots, as described in configuration 4). It should be noted that all of the maps do not represent the same quantity. On the one side, the power-estimate-based beamformers (*i.e.*, FD-DAS, FD-RCap, FD-Pisa) are compared to each other. On the other side, the noise space projection-based beamformer FD-Music requires separate comment (namely, when comparing metrics on a dB scale, as for  $CAR$  or  $Gap$ ).

### A. Simulation results

The four configurations (Fig. 2) are analyzed with the same reconstruction frequency at 5 MHz (*i.e.*, the center frequency of the probe used for the experiments), and with similar pixel spacing of 0.02 mm for the lateral dimension and 0.1 mm for the axial dimension.

#### 1) Simulation, configuration 1: influence of the SNR:

Configuration 1 investigates on a single source the evolution of  $R_{lat}$ ,  $R_{ax}$ ,  $CAR$ , and  $Err_{pos}$  with respect to the SNR, for the four beamformers FD-DAS, FD-RCap with  $\epsilon = 1$  and  $\epsilon = 10$ , FD-Pisa with  $p = 50$  and  $p = 100$ , and FD-Music with  $K_s = 1$  and  $K_s = 3$ . The parameters for the CSM estimation are the same through all of the results for this configuration:  $K = 130$ , with  $p_{ov} = 90\%$ , which corresponds to the optimal parameters described in V-A3. The image quality metrics are shown in Figure 3.

The error positions on Fig. 3d are negligible as they are equal to or smaller than 0.2 mm (*i.e.*, mainly due to one or two pixel errors in the axial dimension, which is consistent with the observations in [50]) for SNRs above -10 dB, except for FD-RCap with  $\epsilon = 1$ . Indeed,  $\epsilon = 1$  corresponds to a relatively small tolerance value for error models. Thus, in the high-noise scenario (SNR below 0 dB), increasing  $\epsilon$  up to 10 reduces the error position to 0.2 mm for SNRs above -10 dB.

In the very-high-noise scenario for SNR = -15 dB, the maps are not reliable for FD-RCap and FD-Music any more, which leads to high position errors (*i.e.*, greater than or equal to 0.7 mm).

For the main lobe size (Fig. 3a, b), the axial width is larger than the lateral width (*i.e.*, about 10-fold), as is commonly seen in passive cavitation imaging [26]. For all beamformers, both of the widths evolved with very similar trends with respect to the SNR. For SNRs above -5 dB, the three adaptive beamformers provide thinner main lobe widths (regardless of the adaptive parameter values), which are almost halved, at least. For FD-RCap, as  $\epsilon$  reflects the tolerance to uncertainties, a larger value of  $\epsilon$  is required for the low SNR, while a smaller value of  $\epsilon$  in the high-noise scenario allows very thin main lobes to be assessed. However, FD-Pisa appears stable even for high  $p$  values, and also FD-Music, with slightly overestimated  $K_s$ , whatever the SNRs. FD-Pisa (for both  $p$  values) provides intermediate performances between FD-DAS and FD-RCap, with the latter appropriately tuned with  $\epsilon$ . Very similar observations are established for  $CAR$  in Fig. 3c (the artifact region is set as in Figure 2, to enclose the FD-DAS upstream side lobes).

This joint analysis on four image quality metrics in the single-source scenario leads to the conclusion that following the guidelines from the literature, the appropriate tuning of the adaptive parameters is not a difficult task for better image quality metrics than for FD-DAS, even for the high-noise scenario (SNR = -5 dB). The two following subsections illustrate the consequences of improved  $R_{ax}$ ,  $R_{lat}$  and  $CAR$ , in terms of enhancement of source localization.

2) *Simulation, configurations 2 and 3: closely spaced sources:* Configurations 2 and 3 are designed as a proof of concept, with the aim to illustrate the superiority of FD adaptive beamformers for the resolution of two closely spaced sources. The four beamformers are considered: FD-DAS, FD-RCap with  $\epsilon = 1$  (*i.e.*, low-noise scenario), FD-Pisa with  $p = 50$ , and FD-Music with  $K_s = 2$  (*i.e.*, two sources). The parameters for the CSM estimation are the same for both configurations:  $K = 130$  with  $p_{ov} = 90\%$ , which corresponds to the optimal parameters described in V-A3. Lateral slices (at  $z = 60$  mm) for configuration 2 and axial slices (at  $x = 5$  mm) for configuration 3, are shown in Fig. 4a, b, respectively.

In both configurations, FD-DAS does not separate the two closely spaced sources, but the adaptive beamformers do. Using FD-Pisa, a slight gap appears between the two sources: lateral  $Gap = 1.0$  dB, and axial  $Gap = 0.4$  dB. FD-RCap provides the best performance among these power-based beamformers, with lateral  $Gap = 6.1$  dB and axial  $Gap = 3.0$  dB. Also, FD-Music clearly detects both sources (as the map units are not the same as the power-based beamformers, we cannot indicate any resolution superiority for FD-Music). As commonly observed in passive cavitation imaging, the axial resolution is worse than the lateral resolution (with source spacing of 0.4 mm on the  $x$  axis, and 3.5 mm on the  $z$  axis). For FD-Pisa and FD-RCap, even if both of these sources nominally transmit the same power, a small difference in the power map estimates is observed for a single reconstruction

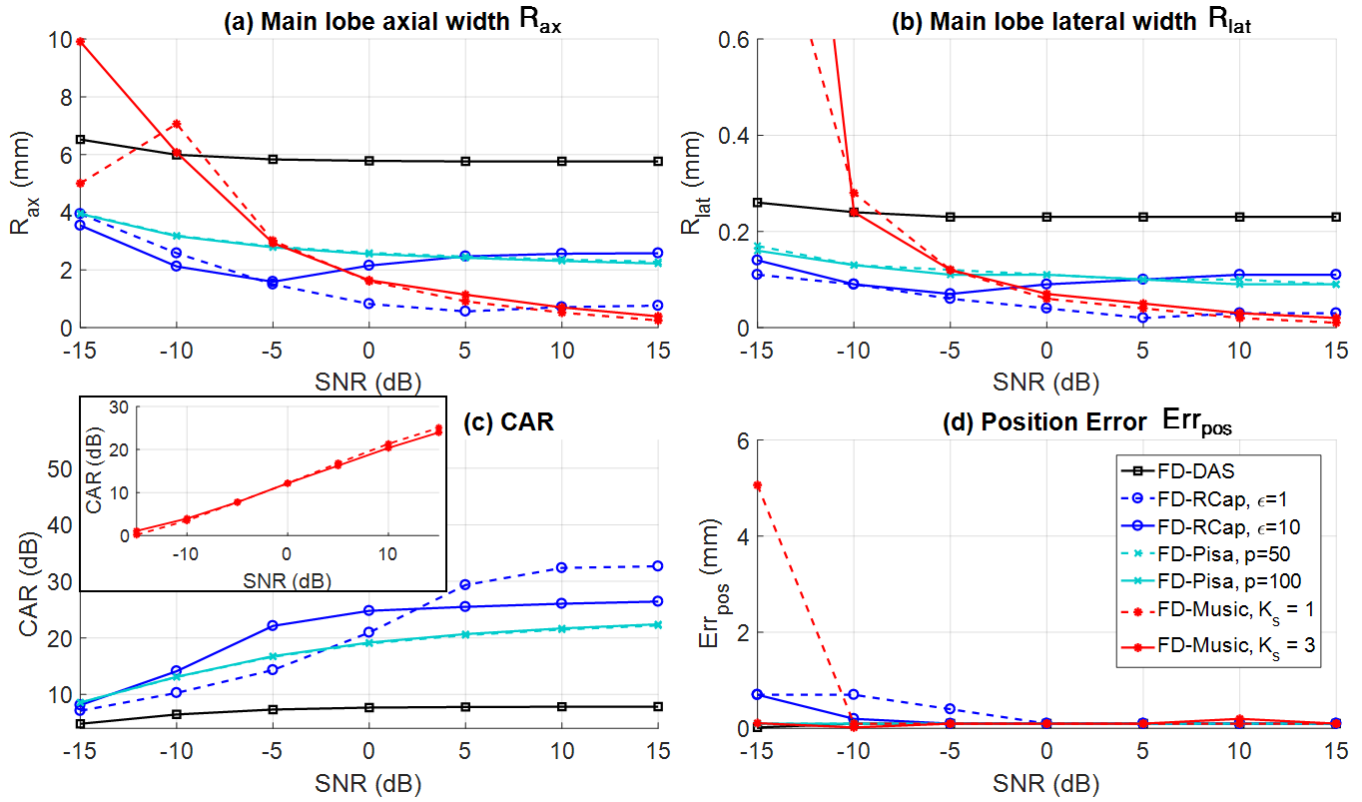


Fig. 3: Performances of the four beamformers, considering the four image quality metrics of (a)  $R_{ax}$ , (b)  $R_{lat}$ , (c)  $CAR$ , and (d)  $Err_{pos}$  versus the SNR, for the single-source simulation (configuration 1). Two adaptive parameters are used for each adaptive beamformer. As FD-Music is not a power-based beamformer, it is set apart for the  $CAR$  plots.

frequency. Indeed, the assumption of white noise is true from a statistical point of view; thus at a single frequency, from one realization to the other, fluctuations of the power spectral density of each source up to 2 dB can be observed. Also, this is potentially due to interference between sources. It should be noted that these configurations are challenging cases that are chosen to highlight the potential of FD adaptive beamformers to resolve two closely spaced sources where the nonadaptive scenario fails.

3) *Simulation, configuration 4: influence of estimation quality of the CSM*: Configuration 4 is designed to highlight the necessity for CSM averaging when using FD adaptive beamformers; namely in a multiple-source scenario. The data for configuration 4 are presented in Figure 5, to provide a strategy to select the parameters for CSM averaging and also to illustrate their influence on the maps. The four beamformers are considered: FD-DAS, FD-RCap with  $\epsilon = 5$  (more sources, more noise, so increased  $\epsilon$  compared to the three previous configurations), FD-Pisa with  $p = 50$ , and FD-Music with  $K_s = 3$  (i.e., three uncorrelated sources).

The strategy to choose the parameters for the CSM averaging is as follows. First, the recording duration  $T_{rec}$  is usually imposed by the hardware capacity or the duration of the observed phenomenon. Moreover,  $K$  is constrained by the condition of inversion of FD-RCap (i.e.,  $K > N$ ; see III-B). So, Fig. 5 first compares the use, or not, of snapshot averaging (i.e.,  $K = 1$  or  $K = 130$ ), and then it illustrates

the effects of  $p_{ov}$  (directly linked to  $T_{snap}$  via Equation (12)). Thus, four cases are considered:  $K = 1$ , i.e., without snapshot averaging (Fig. 5a-d);  $K = 130$  without overlap, which leads to  $T_{snap} = 1.5 \mu s$  (Fig. 5e-h);  $K = 130$  with  $p_{ov} = 50\%$ , which leads to  $T_{snap} = 3.1 \mu s$  (Fig. 5i-l); and  $K = 130$  with  $p_{ov} = 90\%$ , which leads to  $T_{snap} = 15.4 \mu s$  (Fig. 5m-p).

a) *Effects of cross-spectral density matrix averaging on cavitation maps*: Without snapshot averaging ( $K = 1$ ), as in Fig. 5a, FD-DAS detects the strong source, although the sidelobes it generates hide both of the weak sources. In Fig. 5b, FD-RCap does not detect any source, and the map is completely unstable, as  $\hat{\mathbf{M}}(f)$  is ill-conditioned. In Fig. 5c, FD-Pisa only detects the strong source, and misses both of the weak sources, while in Fig. 5d, FD-Music barely detects the strong source, and it has many artifacts. The second row in Fig. 5 corresponds to a number of snapshots where  $K > N$  (without overlap), to ensure the inversion of  $\hat{\mathbf{M}}(f)$  (see III-B). In Fig. 5f, FD-RCap is stabilized. Both of the weak sources emerge for FD-RCap and FD-Pisa (Fig. 5f, g, respectively). However, as  $T_{snap}$  is small, the spectral analysis is degraded, so the main lobe of the strong source is affected (Fig. 5e, g, h), in comparison to the first row (Fig. 5a, c, d). For the third row of Fig. 5, a snapshot overlap is introduced to maintain  $K > N$ , while increasing  $T_{snap}$  with  $p_{ov} = 50\%$  (Fig. 5i-l): the main lobes are sharpened in the axial direction, with respect to the case where  $p_{ov} = 0\%$  (Fig. 5e-h). With  $p_{ov} = 90\%$  (Fig. 5m-p), this trend is enhanced. In Fig. 5m, FD-DAS still fails to

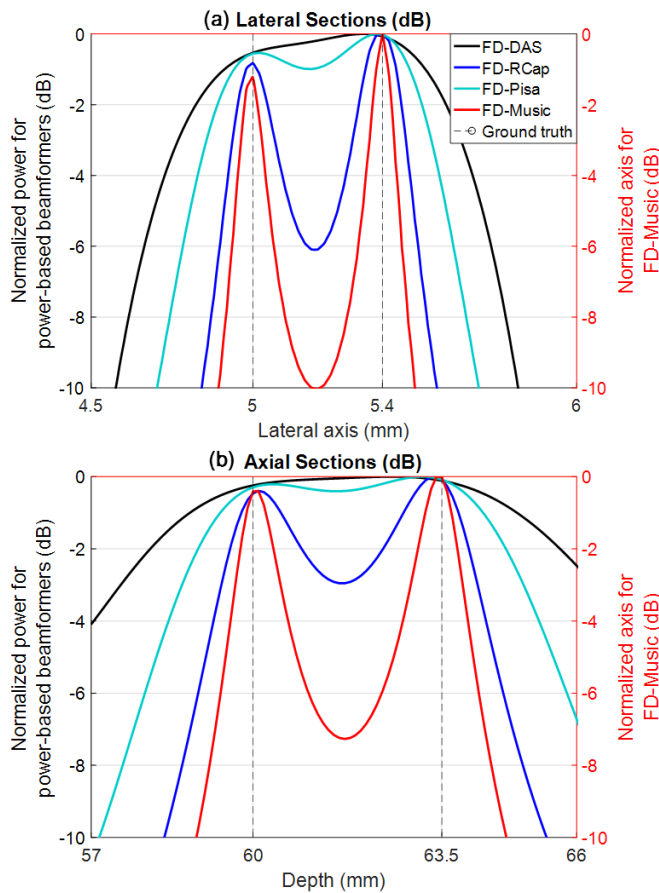


Fig. 4: Lateral slices at  $z = 60$  mm (a) and axial slices at  $x = 5$  mm (b) for two closely spaced sources, as in configurations 2 and 3, respectively. Four beamformers are represented: FD-DAS, FD-RCap with  $\epsilon = 1$ , FD-Pisa with  $p = 50$ , and FD-Music with  $K_s = 2$ . The actual source positions are plotted as the dashed lines. The plots are normalized by their maxima. As FD-Music is not a power-based beamformer, it requires a specific y-axis.

detect both of the weak sources, although all of the sources are detected with the three advanced adaptive beamformers: FD-RCap, FD-Pisa, and FD-Music (Fig. 5n-p). The following analyses compare the beamformers for enhanced weak source detection, with a focus on the set of parameters that gives the best performances (*i.e.*, three detected sources for the three FD adaptive beamformers), such that  $K > N$ ,  $p_{ov} = 90\%$  and  $T_{snap} = 15.4 \mu s$ , which correspond to Fig. 5m-p.

*b) Power-estimate-based beamformer comparison:* The performance is in line with the previous configurations 1, 2, and 3. FD-RCap offers the best performance in terms of resolution (*i.e.*, the thinnest main lobes) and contrast (*i.e.*, the best artifact rejection) (Fig. 5n). However, FD-RCap suffers from the high computational cost (about 104.7 s *versus* 1.3 s for FD-DAS; to compute one map composed of 35,140 pixels). Again, FD-Pisa (Fig. 5o) offers intermediate performance between FD-DAS (Fig. 5m) and FD-RCap (Fig. 5n), although with an execution time of 1.4 s, so similar to FD-DAS. The two weak sources are not distinguished with FD-RCap either,

although they are clearly revealed in comparison to FD-DAS.

*c) Influence of  $K_s$  on FD-Music:* The four maps obtained with FD-Music are shown in Fig. 6, as:  $K_s = 1$  (Fig. 6d) (with underestimated size of the signal subspace  $\mathbf{E}_s$ );  $K_s = 3$  (Fig. 6e) (with correctly estimated size of  $\mathbf{E}_s$ ); and  $K_s = 5$  (Fig. 6f) or  $K_s = 10$  (Fig. 6g) (with overestimated size of  $\mathbf{E}_s$ ). The EVs are shown in Fig. 6h. When the real size of the signal subspace is underestimated, some of the sources are missing, as in Fig. 6d. When  $K_s$  is the same as the effective size of the signal subspace, all of the sources are correctly detected (Fig. 6e) with the best performance metrics: the *Gap* between the two weak sources is 8 dB. However, when the effective size of the signal subspace is overestimated ( $K_s = 5$  Fig. 6f, or  $K_s = 10$  (Fig. 6g), the maps are degraded (*i.e.*, with artifacts and main lobe extensions), with decreased localization performance (*Gap* between the two weak sources of 5.7 dB and 4.6 dB, respectively).

When the real size of the signal subspace is not known, this can be estimated considering the EV plot (Fig. 6h) [47]. A simple trial-and-error approach can be as follows. The sources are decomposed on the first EVs, then they are followed by a plateau for the EVs due to a spatially decorrelated noise; see Equation (10) or [31]. Thus, the plateaus (*e.g.*, EV 2 and 3) or bends (*e.g.*, EV 2 or 5) observed on the first eigenvalues before a broad plateau or a slow decrease in EV, are potentially the values for which the signal space and the noise space separate. As underestimation of the signal space can miss the source detection (Fig. 6d), it is better to slightly overestimate this space (Fig. 6f), although not by too much, to avoid loss of resolution or contrast (Fig. 6g). This approach is used to tune  $K_s$  for the experimental results. More sophisticated methods can be investigated, although they are application dependent (*e.g.*, see the exhaustive comparison of automated singular value thresholding for ultrasound Doppler imaging in [51]). Finally, whatever  $K_s$  is, the FD-Music running time is 1.4 s, and so similar to that of FD-DAS.

*d) Joint qualitative comparison of the four beamformers:* Figure 6 qualitatively compares the beamformers for enhanced weak source detection. The maps from Fig. 5m-p are shown in Fig. 6a-c, e with the specific dynamic ranges, from the lowest intensity in their own background, up to 0 dB. The displaying of these maps with different dynamic ranges is suitable to compare the localizations of the different sources. Indeed, it can be seen that FD-Music (Fig. 6e) provides a very smoothed background with respect to FD-RCap (Fig. 6b) and FD-Pisa (Fig. 6c). In this way, FD-Music is expected to be an excellent tool to distinguish sources from artifacts.

## B. Experimental results

*1) Single wire configuration:* Fig. 7 shows the maps obtained with FD-DAS (Fig. 7a), FD-RCap (Fig. 7b) ( $\epsilon = 10$ ), FD-Pisa (Fig. 7c) ( $p = 30$ ), and FD-Music (Fig. 7d) ( $K_s = 3$ ), with the overlaying optical images. The configuration shown in Fig. 7 to illustrate the performances of the beamformers for the real cavitation events corresponds to an oxygen concentration of 2.5 mg/L, a measured cavitation threshold of 5 MPa (*i.e.*, a little smaller than the peak-negative pressure at the

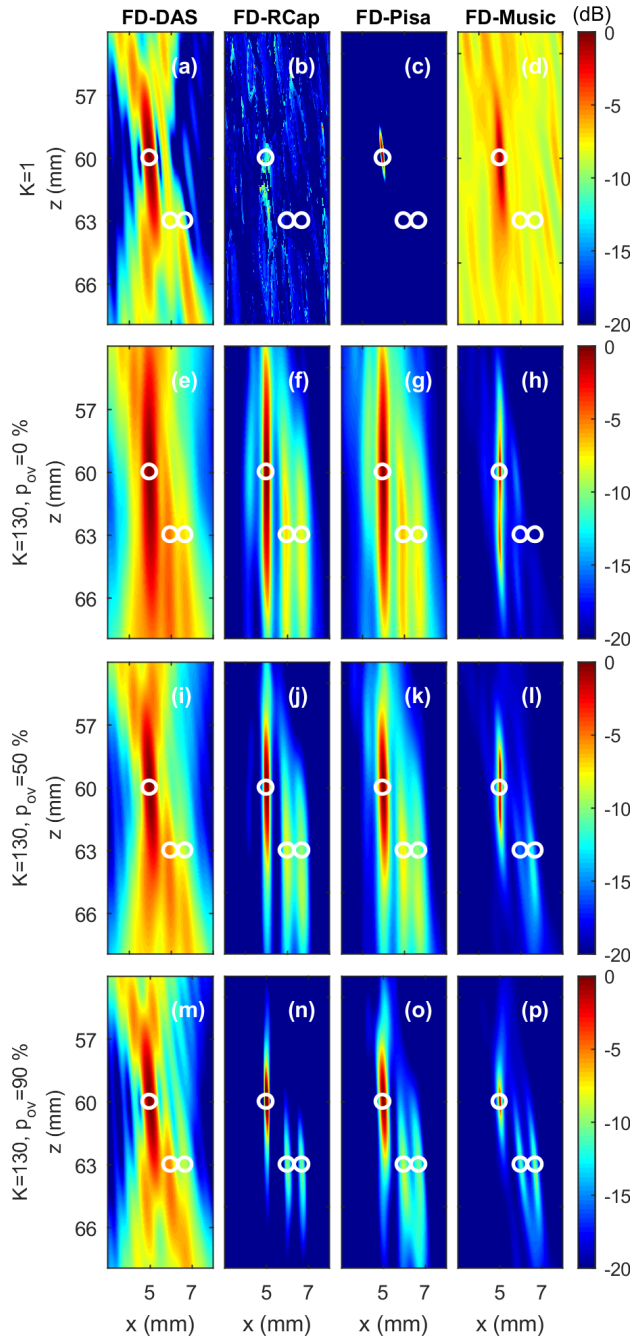


Fig. 5: Influence of the cross-spectral density matrix estimation on the cavitation maps obtained with the four different beamformers, in simulation. The maps are reconstructed at 5 MHz. Each column corresponds to a beamformer, as indicated. For FD-RCap,  $\epsilon = 5$ , for FD-Pisa,  $p = 50$ , and for FD-Music,  $K_s = 3$ . Different parameters for CSM estimation are tested for each row: (a-d)  $K = 1$  (i.e., without snapshot averaging); (e-h)  $K = 130$  without overlap, which leads to  $T_{snap} = 1.5 \mu s$ ; (i-l)  $K = 130$  with  $p_{ov} = 50 \%$ , which leads to  $T_{snap} = 3.1 \mu s$ ; and (m-p)  $K = 130$   $p_{ov} = 90 \%$ , which leads to  $T_{snap} = 15.4 \mu s$ . The actual bubble positions are the centers of the white circles. All of the maps are normalized and displayed with the same dynamic range, of 20 dB.

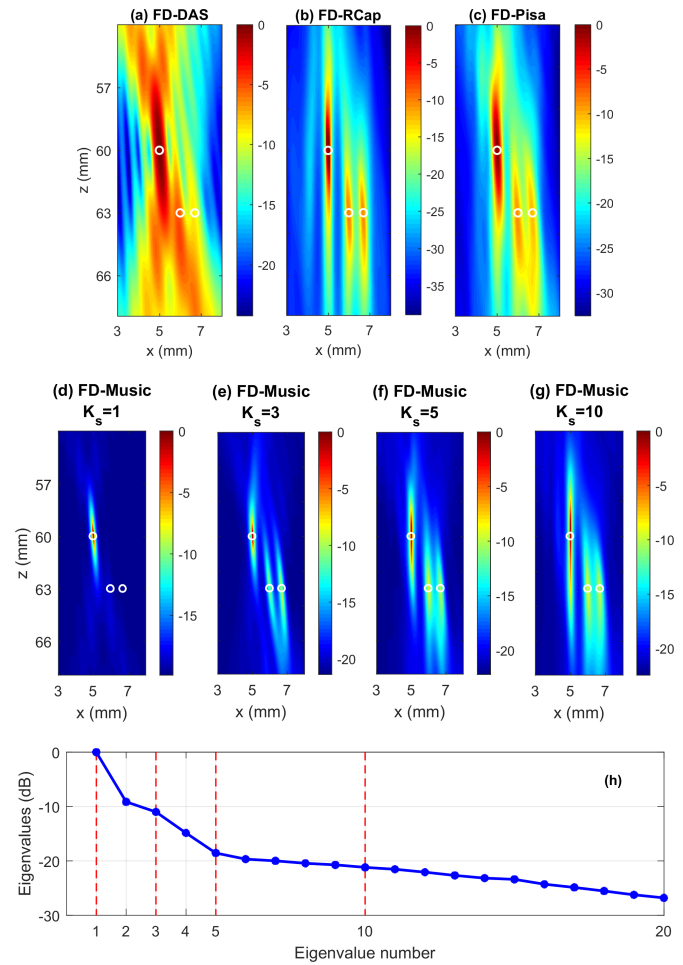


Fig. 6: Beamformer comparisons in simulation. The maps are reconstructed at 5 MHz. All of the maps are reconstructed with  $K = 130$  and  $p_{ov} = 90 \%$  for CSM averaging. The first row corresponds to the power-estimate-based beamformers with optimal settings for FD-DAS (a), FD-RCap with  $\epsilon = 5$  (b), and FD-Pisa with  $p = 50$  (c). The second row corresponds to the maps obtained for FD-Music, with  $K_s = 1$  (d),  $K_s = 3$  (e),  $K_s = 5$  (f), and  $K_s = 10$  (g). (h) The normalized eigenvalues are plotted logarithmically. The actual source positions are the centers of the white circles. (a, b, c, e) correspond to the maps of Fig. 5m, n, o, p, respectively, with different dynamic ranges for the displays. The maps are normalized and displayed with their own full dynamic ranges: from the lowest intensity in their own background, up to 0 dB.

focus of 7.2 MPa) in the middle of the shot at  $t = 10$  ms. Only a few bubbles are detected with the optical camera, and they are all located on the wire, in the small area that is 1-mm long and is circled by the white/ red dashed lines in Fig. 7. A reconstruction frequency of 4.7 MHz was chosen, close to the central frequency of the probe, to ensure that only inertial cavitation activity is mapped, and not the harmonic contribution that results from nonlinear distortion of the acoustic pulse due to the transducer itself, and potentially scattered by the wire [26]. It should be noted that during the shot, the wire was pushed away from the transducer a little, with the maximum



axial displacement of 0.3 mm. This wire displacement was not seen in the absence of cavitation, and results from the radiation pressure on the bubbles. For all of the maps, the beamforming methods succeed in localizing the cavitation activity with a single main lobe that coincides with the area where the bubbles are detected by the optical camera. This remains true over the entire shot, with the main lobe following the slight displacement of the wire. The characteristic hour-glass shape of FD-DAS (Fig. 7a), which has no physical meaning, is absent from the adaptive approach maps. FD-RCap (Fig. 7b) shows a smooth, oval lobe, as already seen in the literature for its time-domain version [15]. This is also the case for FD-Pisa (Fig. 7c), while the main lobe of FD-Music (Fig. 7d) is less homogeneous and appears to agglomerate several small lobes. The localization performance statistics calculated on the 100 maps reconstructed during the 20 ms of shooting are shown in Fig. 7e for axial resolution  $R_{ax}$ , Fig. 7f for lateral resolution  $R_{lat}$ , and Fig. 7g for  $CAR$ . For the power-estimate-based beamformers, while the lateral resolution  $R_{lat}$  is almost identical from one method to another, the axial resolution  $R_{ax}$  is clearly better for the adaptive approaches, with an average of 7.5 mm for FD-DAS, *versus* only 3.7 mm for FD-RCap and 4.6 mm for FD-Pisa. The  $CAR$  (Fig. 7g) is also greatly improved with the adaptive methods. Once again, FD-Music cannot be directly quantitatively compared to the power-estimate-based beamformers, as it does not represent the same quantities. Nevertheless, the qualitative aspect for FD-Music described above is reflected in the quantitative metrics with relative considerations between the metrics. Indeed, the  $R_{lat}$  metrics for FD-Music are very close to those of FD-DAS (less than 0.08 mm difference, for mean values), while  $R_{ax}$  and  $CAR$  are clearly enhanced for FD-Music compared to FD-DAS, with an enhancement of more than 5 mm and 5 dB for the mean values, respectively.

2) *Dual wire configuration*: In the second configuration, which included the two wires, the oxygen concentration was increased to 6.2 mg/L, which reduces the cavitation threshold to below 0.7 MPa while keeping the excitation HIFU signal unchanged. This enabled the triggering of a second cavitation spot on the second wire that was placed downstream from the focus of the HIFU transducer. The maximum displacements of the wires due to the radiation pressure on the cavitation bubbles were 1.6 mm for the upstream wire and 0.6 mm for the downstream wire. These values are higher than for the first configuration, which is attributed to more sustained cavitation activity detected on the wires in optical images in the circled regions in Fig. 8a-d. As anticipated, the cavitation activity observed in the optical images appears to be more intense for the upstream wire, which was placed closer to the focus of the HIFU transducer than the downstream wire. Fig. 8 shows the representative maps of two cavitation sources on the two separate wires for FD-DAS (Fig. 8a), FD-RCap (Fig. 8b) ( $\epsilon = 10$ ), FD-Pisa (Fig. 8c) ( $p = 30$ ), and FD-Music (Fig. 8d) ( $K_s = 4$ ) during the first part of the shot, at  $t = 4$  ms. The reconstruction frequency of 4.7 MHz was chosen. For all of the maps, both cavitating regions correspond to the two main lobes of the maps, which are centered on each wire at about 59 mm in depth, where the cavitation activity was

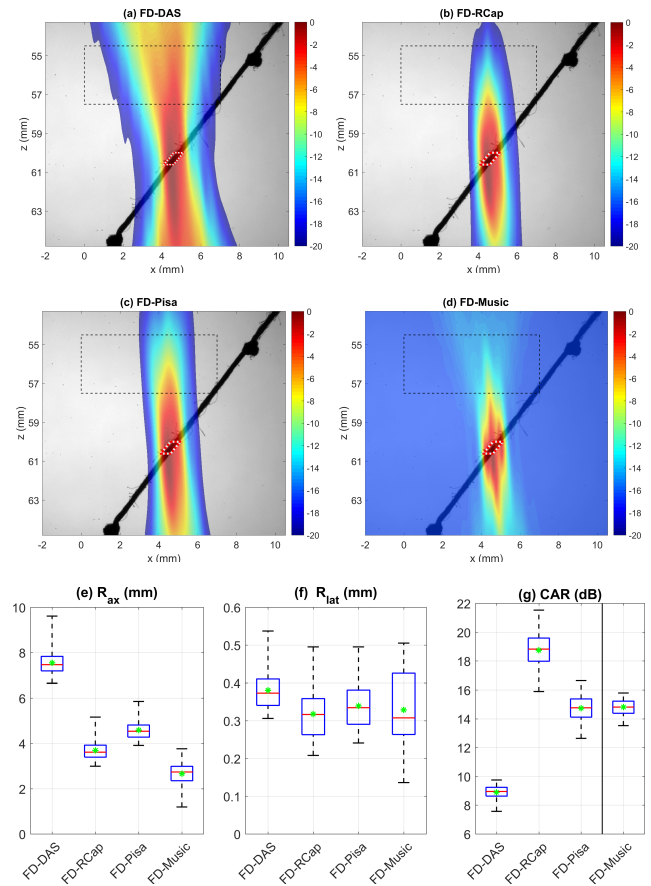


Fig. 7: Representative experimental localization maps for a single cavitation spot on a wire immersed in degassed and filtered water, for FD-DAS (a), FD-RCap (b) ( $\epsilon = 10$ ), FD-Pisa (c) ( $p = 30$ ), and FD-Music (d) ( $K_s = 3$ ), in the middle of the HIFU shot (10 ms). The HIFU transducer is on the right in the images. The maps are reconstructed at 4.7 MHz. The bubbles detected with the optical camera are located within the white/ red dashed lines. The dotted rectangle indicates the artifact region for the  $CAR$  measurement, with respect to the maximum amplitude of the map. The performance statistics are calculated on the 100 maps reconstructed during the 20 ms of shooting, and are shown for axial resolution (e), lateral resolution (f), and  $CAR$  (g), with the median in red, the mean in green, the first and third quartiles in blue, and the total extent in black.

detected optically. For the power-estimate-based beamformers, the source on the downstream wire is weaker. For FD-DAS, its intensity hardly reaches the level in the characteristic hourglass-shaped artifact from the strong source: this might have been hidden in another geometric configuration, such as in Fig. 5. For the adaptive beamformers FD-RCap and FD-Pisa, the better resolution allows the detection of the weak source more clearly. FD-Music gives a very different map, with two clear main lobes at the positions of the cavitating regions, whereby both emerge very clearly from a very smooth and homogeneous background, which is similar to what was noted in the simulation data. Again, the lobes appear to have a more complex internal structure than for the other beamformers.

3) *Spatio-temporal monitoring*: The projected maximum map values on the  $x$ -axis *versus* time are shown for each algorithm in Fig. 8e-h. This representation is interesting for measuring the ability of each method to bring out the main lobe of a source with respect to the background map, and to ensure the spatio-temporal monitoring of the cavitation with respect to the  $x$ -axis (similar to the spatio-temporal tracking in [48]). The evolution of the local maxima of the maps was correlated with the displacement of the wires on which the cavitation bubbles were created (Fig. 8e-h, two dashed curved lines). In all of the cases, two trails are observed. The right-wing trace (at around  $x = 4$  mm) was very energetic for the first half of the shot, and corresponds to the main lobe centered on the upstream wire (wire #1). The left-wing trace (at around  $x = 1$  mm) was less energetic, and corresponds to the lobe centered on the downstream wire (wire #2). However, for FD-DAS (Fig. 8e) up to about  $t = 10$  ms, this left-wing trace hardly emerges from artifacts from the more energetic right trace, while the adaptive methods FD-RCap (Fig. 8f) and FD-Pisa (Fig. 8g) efficiently reject these artifacts, and show clear identification of the left trace, thus enhancing the spatio-temporal monitoring of the cavitation activity. For FD-Music (Fig. 8h), the cavitation cloud on wire #2 is raised to almost the same level as the cavitation on wire #1. Again, this observation is consistent with the lack of preservation of the relative amplitudes of the sources using FD-Music, and makes it possible to highlight sources with little power. The absence of cavitation is detected on wire #1 beyond 10 ms for all of the systems here (Fig. 8e-h). The growth of the cavitation bubbles involves the recruitment of gas pockets trapped on the wire surface, as well as gas dissolved in the surrounding medium. As soon as larger-sized bubbles are obtained, they are pushed away by the radiation force of the acoustic wave, which degasses the wire and its surroundings, so that cavitation on the wire would require a higher magnitude of negative pressure. Finally, particularly with the adaptive beamformers, passive ultrasound imaging allows good spatio-temporal monitoring of complex scenarios of cavitation events. This successfully detects the extinction of the first cavitation spot around  $t = 10$  ms, and the persistence of the second cavitation spot through the whole burst. The specificity of FD-Music can also be emphasized, as it unambiguously highlights the weak source from the beginning to the end of the burst.

## VI. DISCUSSION

In this paper, the concept of robust CSM estimation is introduced and its benefits for passive cavitation imaging are illustrated for four beamformers, through both simulation and experiments. In this way, we demonstrate that adaptive beamformers, and the enhanced image quality they provide, are compatible with low computational cost and frequency selectivity provided by FD approaches.

1) *Robust cross-spectral density matrix estimation is required*: Fig. 5 shows the need for robust CSM estimation to develop adaptive beamformers. As beamformer theory is derived from the theoretical CSM  $\mathbf{M}(f)$  expressed as an expected value  $E\langle \cdot \rangle$ , it is not surprising that averaging observations

highly impacts upon the performance of the beamformers: this ensures the required matrix inversion for FD-RCap and a reliable EVD for the CSM for FD-Pisa and FD-Music. In this way, the maps are stabilized and the sources are correctly detected (Fig. 5). This result is well known in the array-processing field, and in particular for DOA estimation [21], [34]. Averaging with snapshots is often referred to as the natural CSM estimation, although alternative or joint strategies can be considered, such as diagonal loading [50] and spatial smoothing [18]. Note that for the robust Capon beamformer in the time domain [15], the spatial covariance matrix is estimated from several observations, each of which corresponding to one acquisition time sample.

2) *Power-estimate-based beamformers comparison*: The CSM is the basis from which many beamformers can be derived [31], [32], [35]. Three beamformers are chosen here to illustrate the benefits of robust CSM estimation, and a preliminary comparison can be drawn between the power-estimate-based ones. The nonadaptive FD-DAS approach offers equivalent performance to [26] (in the special case of FD-DAS for which the CSM estimation is based on a single observation). FD-DAS is used here as a nonadaptive reference method, as useful for comparing more advanced approaches. The FD-RCap method improves source localization in terms of resolution and contrast, as already demonstrated in its time-domain version (TD-RCap) [15]. FD-RCap is therefore an excellent candidate for high-resolution localization of cavitation. Its main disadvantage is the computational overload that is required for the pixel-wise matrix inversion (see Equation (20)), with respect to FD-DAS. Note that for TD-RCap [15], the full set of the received signals  $\mathbf{y}(t)$  has to be delayed for each pixel, and then used to construct the spatial covariance matrix, which is highly time consuming. In the FD, the CSM is obtained directly from raw signals and efficiently steered with fast matrix products. In this way, although the optimization operation with the inversion of the matrix remains with a huge amount of calculus, FD-RCap is a relative fast version for a given frequency of TD-RCap (similar comparison as in [27]). As a point of comparison, and to give an order of magnitude to the time that can be saved, for the maps that are composed of 35,140 pixels, the low-complexity FD beamformers (*i.e.*, FD-DAS with or without snapshot averaging, or FD-Pisa) run in the order of a few seconds, whereas the higher complexity FD beamformer (FD-RCap) requires one or a few minutes, and the time-domain beamformers (TD-DAS, TD-RCap) easily exceed an hour, mainly due to the very time consuming delaying step over 200 s of the integration time. Alternatively, the FD-Pisa, as the Pisarenko class beamforming, offers intermediate performance between FD-DAS and FD-RCap for a computing load equivalent to FD-DAS. Finally, from a statistical point of view, the CSM formalism developed for power-estimate-based beamformers addresses the issue of quantitative power estimation of cavitation, which is an active and challenging field of research in terms of both numerical [26] and instrumentation [52] aspects.

3) *Focus on FD-Music, the nonpower-estimate-based approach*: Apart from the power-estimate-based beamformers, FD-Music is well known as a high-resolution localization



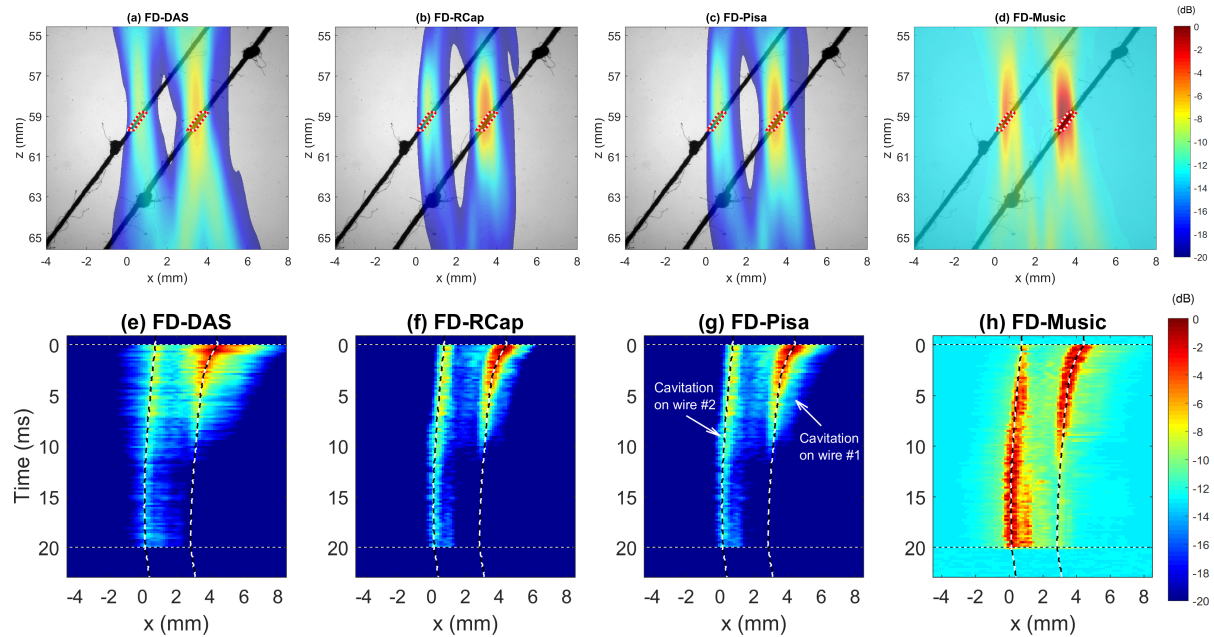


Fig. 8: Representative experimental localization maps for the two cavitation spots on the two wires immersed in the filtered water, for FD-DAS (a), FD-RCap (b) ( $\epsilon = 10$ ), FD-Pisa (c) ( $p = 30$ ), and FD-Music (d) ( $K_s = 4$ ), at  $t=4$  ms. The HIFU transducer is on the right of the images. The maps are reconstructed at 4.7 MHz. The bubbles detected with the optical camera are located within the white/red dashed lines. Projected maximum map values on the  $x$ -axis *versus* time is shown for each map, respectively (e-h). The projection considers the depth range 54 mm to 68 mm. The vertical dashed lines are the positions of the wire centers throughout the shot.

method [31]. It has been demonstrated that FD-Music highlights the presence of sources, but does not provide a pixel-wise power estimate: absolute quantification or relative quantification of the power of a source is not possible. When monitoring cavitation, weak power sources have to be detected very early in their development if they are located in a non-targeted area, to avoid undesirable effects like healthy tissue ablation. The simulation results in Fig. 6 and the experimental results in Fig. 8 show that FD-Music efficiently highlights weak sources. A specific aspect of this beamformer is that the background is smoother than that of the power-estimate-based beamformers (*i.e.*, FD-DAS, FD-RCap, FD-Pisa), and therefore weak power sources are clearly identifiable. FD-Music is thus an interesting tool for localization of weak power sources, a situation where power-estimate-based beamformers are not appropriate. Furthermore, FD-Music has a running time similar to FD-DAS. Finally, unlike  $\epsilon$  for FD-RCap or  $p$  for FD-Pisa, which are chosen empirically [15], [43], the adaptive parameter  $K_s$  of FD-Music can be adaptively chosen in line with the number of sources. For this experimental study, a very simple trial-and-error approach is used, as described in simulation, as a trade-off between the correct number of expected sources and the image quality metrics. However, the deduced  $K_s$  values are slightly above our expectations (*i.e.*,  $K_s = 3$  for a single spot;  $K_s = 4$  for two spots). This would suggest that a single bubble spot should instead be looked upon as several sources. Note also that for the sake of simplicity, we consider a constant value for  $K_s$  that is relevant through the whole HIFU shot duration. Automated and more sophisticated

EV thresholding can be investigated [47], [51].

#### 4) Methodological perspectives :

The proposed formalism is based on second-order statistics of CSM and its subsequent assumptions of ergodic and stationary signals. This assumption is often assessed with respect to acquisition duration. This preliminary study illustrates that in some simple cavitation scenarios, these assumptions are verified, although further investigations are needed. For example, a higher-order statistics beamformer has been proposed for passive cavitation imaging in the time domain [53] that overcomes the TD-RCap resolution. Also, the assumption for uncorrelated sources should be investigated for extended bubble clouds.

These results demonstrate that single-frequency imaging provides accurate localization of cavitation, even if its spectral signature is broad. In the passive cavitation imaging literature [26], [27], [54], due to the frequency selectivity of FD approaches, single-frequency imaging is often considered to enable discrimination of stable cavitation from inertial cavitation. However, composite multifrequency images can also be considered by superimposing maps at different frequencies (*i.e.*, summing) to further reject incoherent noise on maps from one frequency bin to another [26], [27], [54]. Note that the computing load increases almost proportionally with the number of reconstructed bins [27]. Moreover, as the beamwidth is frequency dependent [26], constant-beamwidth approaches can be considered for better consistency of acoustic maps, even for adaptive beamformers [55]. A specific aspect to adaptive beamformers would be that the adaptive parameters can be

frequency dependent, and as the error uncertainties might not be constant over the full bandwidth of the probe, this would potentially influence  $\epsilon$  and  $p$  [43], [55].

This paper focuses on the enhancement of detection and localization of possibly simultaneous cavitation events, but to go further, the integration of maps over frequencies from power-estimate-based beamformers might provide access to power quantification within a given band, as in [21], [26], [44]. Nevertheless, this requires further challenging considerations, which include fine numerical scaling [26], [56], and also the appropriate calibration and acoustic modeling [15], [52], which are beyond the scope of this paper.

Cavitation is a three-dimensional physical phenomenon. Imaging this with linear arrays limits its localization into a two-dimensional plane. The subsequent issues of this aspect are that off-plane cavitation events might not only be missed, but might also generate artifacts on the two-dimensional maps. On the one hand, the potential for adaptive beamformers to remove such artifacts has to be investigated. On the other hand, the CSM-derived beamformers are good candidates to deal with three-dimensional imaging. As these are intrinsically formalised in the FD, they are very computationally efficient [27]. This property is essential to efficiently process the large volumes of data involved in three-dimensional imaging, for which more than 1000 sensors can be used [57].

## VII. CONCLUSIONS

This study has shown the potential of CSM-derived beamformers for spatio-temporal monitoring of cavitation. These provide: (1) high quality maps in (2) shorter computational times than needed for time-domain approaches, and with (3) direct frequency selectivity. These three attractive properties are gathered within the proposed CSM-derived formalism, which is therefore expected to be a powerful tool to develop passive ultrasound imaging. In the long term, these CSM-derived beamformers can be expected to contribute to the enhancement of spatio-temporal regulation of cavitation for therapeutic applications [54].

## REFERENCES

- [1] J.-M. Escoffre and A. Bouakaz, *Therapeutic ultrasound*. Springer, 2015.
- [2] M. Wan, Y. Feng, and G. ter Haar, *Cavitation in biomedicine*. Springer, 2015.
- [3] V. A. Khokhlova, M. R. Bailey, J. A. Reed, B. W. Cunitz, P. J. Kaczowski, and L. A. Crum, "Effects of nonlinear propagation, cavitation, and boiling in lesion formation by high intensity focused ultrasound in a gel phantom," *J. Acoust. Soc. Am.*, vol. 119, no. 3, p. 1834, Feb 2006.
- [4] C. H. Farny, R. G. Holt, and R. A. Roy, "The correlation between bubble-enhanced HIFU heating and cavitation power," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 1, pp. 175–184, Jan 2010.
- [5] Y. A. Pishchalnikov, O. A. Sapozhnikov, M. R. Bailey, J. C. Williams, Jr., R. O. Cleveland, T. Colonius, L. A. Crum, A. P. Evan, and J. A. McAtteer, "Cavitation bubble cluster activity in the breakage of kidney stones by lithotripter shock waves," *Journal of endourology / Endourological Society*, vol. 17, no. 7, p. 435, Sep 2003.
- [6] P. Qin, T. Han, A. C. H. Yu, and L. Xu, "Mechanistic understanding the bioeffects of ultrasound-driven microbubbles to enhance macromolecule delivery," *J. Control. Release*, vol. 272, pp. 169–181, Feb 2018.
- [7] C. D. Arvanitis, M. S. Livingstone, and N. McDannold, "Combined ultrasound and MR imaging to guide focused ultrasound therapies in the brain," *Phys. Med. Biol.*, vol. 58, no. 14, pp. 4749–4761, Jul 2013.
- [8] K. B. Bader, M. J. Gruber, and C. K. Holland, "Shaken and stirred: mechanisms of ultrasound-enhanced thrombolysis," *Ultrasound Med. Biol.*, vol. 41, no. 1, pp. 187–196, Jan 2015.
- [9] Y. Lu, J. Wang, R. Huang, G. Chen, L. Zhong, S. Shen, C. Zhang, X. Li, S. Cao, W. Liao, Y. Liao, and J. Bin, "Microbubble-mediated sonothrombolysis improves outcome after thrombotic microembolism-induced acute ischemic stroke," *Stroke*, vol. 47, no. 5, pp. 1344–1353, May 2016.
- [10] V. A. Khokhlova, J. B. Fowlkes, W. W. Roberts, G. R. Schade, Z. Xu, T. D. Khokhlova, T. L. Hall, A. D. Maxwell, Y.-N. Wang, and C. A. Cain, "Histotripsy methods in mechanical disintegration of tissue: towards clinical applications," *Int. J. Hyperthermia*, vol. 31, no. 2, pp. 145–162, Mar 2015.
- [11] 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.
- [12] M. Gyöngy and C. C. Coussios, "Passive spatial mapping of inertial cavitation during HIFU exposure," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 1, pp. 48–56, 2010.
- [13] S. J. Norton and I. Won, "Time exposure acoustics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1337–1343, 2000.
- [14] S. J. Norton, B. J. Carr, and A. J. Witten, "Passive imaging of underground acoustic sources," *The Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 2840–2847, 2006.
- [15] C. Coviello, R. Kozick, J. Choi, M. Gyöngy, C. Jensen, P. P. Smith, and C. C. Coussios, "Passive acoustic mapping utilizing optimal beamforming in ultrasound therapy monitoring," *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 2573–2585, 2015.
- [16] P. Boulos, F. Varray, A. Poizat, A. Ramalli, B. Gilles, J.-C. Bera, and C. Cachard, "Weighting the passive acoustic mapping technique with the phase coherence factor for passive ultrasound imaging of ultrasound-induced cavitation," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 65, no. 12, pp. 2301–2310, 2018.
- [17] M. Polichetti, F. Varray, J.-C. Béra, C. Cachard, and B. Nicolas, "Advanced beamforming techniques for passive imaging of stable and inertial cavitation," in *IEEE International Ultrasonics Symposium*, 2018.
- [18] J. Synnevag, A. Austeng, and S. Holm, "Adaptive beamforming applied to medical ultrasound imaging," *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, vol. 54, no. 8, p. 1606, 2007.
- [19] J. Camacho, M. Parrilla, and C. Fritsch, "Phase coherence imaging," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 56, no. 5, 2009.
- [20] M. Polichetti, F. Varray, J.-C. Béra, C. Cachard, and B. Nicolas, "A nonlinear beamformer based on p-th root compression - Application to plane wave ultrasound imaging," *Applied Sciences*, vol. 8, no. 4, p. 599, 2018.
- [21] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE transactions on signal processing*, vol. 51, no. 7, pp. 1702–1715, 2003.
- [22] P. Stoica, Z. Wang, and J. Li, "Robust Capon beamforming," *IEEE Signal Process. Lett.*, vol. 10, no. 6, pp. 172–175, Jun 2003.
- [23] S. Lu, A. Shi, B. Jing, X. Du, and M. Wan, "Real-time monitoring of controllable cavitation erosion in a vessel phantom with passive acoustic mapping," *Ultrason. Sonochem.*, vol. 39, pp. 291–300, Nov 2017.
- [24] S. Lu, H. Hu, X. Yu, J. Long, B. Jing, Y. Zong, and M. Wan, "Passive acoustic mapping of cavitation using eigenspace-based robust Capon beamformer in ultrasound therapy," *Ultrason. Sonochem.*, vol. 41, pp. 670–679, Mar 2018.
- [25] K. J. Haworth, T. D. Mast, K. Radhakrishnan, M. T. Burgess, J. A. Kopechek, S.-L. Huang, D. D. McPherson, and C. K. Holland, "Passive imaging with pulsed ultrasound insonations," *J. Acoust. Soc. Am.*, vol. 132, no. 1, p. 544, Jul 2012.
- [26] K. J. Haworth, K. B. Bader, K. T. Rich, C. K. Holland, and T. D. Mast, "Quantitative frequency-domain passive cavitation imaging," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 64, no. 1, pp. 177–191, 2017.
- [27] C. D. Arvanitis, C. Crake, N. McDannold, and G. T. Clement, "Passive acoustic mapping with the angular spectrum method," *IEEE transactions on medical imaging*, vol. 36, no. 4, pp. 983–993, 2017.
- [28] J. McLaughlan, I. Rivens, T. Leighton, and G. ter Haar, "A study of bubble activity generated in ex vivo tissue by high intensity focused ultrasound," *Ultrasound in medicine & biology*, vol. 36, no. 8, pp. 1327–1344, 2010.
- [29] 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," *The*

- Journal of the Acoustical Society of America*, vol. 140, no. 1, pp. 741–754, 2016.
- [30] C. R. Jensen, R. W. Ritchie, M. Gyöngy, J. R. 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, 2012.
  - [31] H. Krim and M. Viberg, “Two decades of array signal processing research: the parametric approach,” *IEEE Signal Process Mag.*, vol. 13, no. 4, pp. 67–94, Jul 1996.
  - [32] B. D. Van Veen and K. M. Buckley, “Beamforming: a versatile approach to spatial filtering,” *IEEE assp magazine*, vol. 5, no. 2, pp. 4–24, 1988.
  - [33] M. R. Azimi-Sadjadi, A. Pezeshki, and N. Roseveare, “Wideband DOA estimation algorithms for multiple moving sources using unattended acoustic sensors,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 4, pp. 1585–1599, Oct 2008.
  - [34] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.
  - [35] R. Merino-Martinez *et al*, “A review of acoustic imaging methods using phased microphone arrays,” *CEAS Aeronaut. J.*, vol. 10, no. 1, pp. 197–230, Mar 2019.
  - [36] S. A. Vorobyov, “Principles of minimum variance robust adaptive beamforming design,” *Signal Process.*, vol. 93, no. 12, pp. 3264–3277, Dec 2013.
  - [37] K. B. Bader, K. J. Haworth, A. D. Maxwell, and C. K. Holland, “Post hoc analysis of passive cavitation imaging for classification of histotripsy-induced liquefaction *in vitro*,” *IEEE Trans. Med. Imaging*, vol. 37, no. 1, pp. 106–115, Jan 2018.
  - [38] P. Stoica, J. Li, and X. Tan, “On spatial power spectrum and signal estimation using the pisarenko framework,” *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5109–5119, 2008.
  - [39] H. Mermoz, “Imagerie, corrélation et modèles,” in *Annales des télécommunications*, vol. 31, no. 1-2. Springer, 1976, pp. 17–36.
  - [40] P. Stoica and R. Moses, *Spectral analysis of signals*. Pearson Prentice Hall Upper Saddle River, NJ, 2005, ch. Spatial Methods.
  - [41] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.
  - [42] V. Pisarenko, “On the estimation of spectra by means of non-linear functions of the covariance matrix,” *Geophysical Journal International*, vol. 28, no. 5, pp. 511–531, 1972.
  - [43] R. P. Dougherty, “Functional Beamforming,” in *Berlin Beamforming Conference*, 2014.
  - [44] V. Baron, A. Finez, and B. Nicolas, “Numerical and experimental assessment of functional beamforming for source quantification,” in *Berlin Beamforming Conference*, 2018.
  - [45] G. Bienvenu and L. Kopp, “Optimality of high resolution array processing using the eigensystem approach,” *IEEE Transactions on acoustics, speech, and signal processing*, vol. 31, no. 5, pp. 1235–1248, 1983.
  - [46] S. Marcos *et al.*, *Les méthodes à haute résolution*, 1998, vol. 42, ch. Sélection d’ordre de modèle pour les méthodes HR.
  - [47] K. M. Wong, Q.-T. Zhang, J. P. Reilly, and P. Yip, “On information theoretic criteria for determining the number of signals in high resolution array processing,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 11, pp. 1959–1971, 1990.
  - [48] 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,” *Physics in Medicine & Biology*, vol. 59, no. 17, p. 4861, 2014.
  - [49] E. A. Neppiras, “Measurement of acoustic cavitation,” *IEEE Transactions on Sonics and Ultrasonics*, vol. 15, no. 2, pp. 81–88, Apr 1968.
  - [50] S. H. Abadi, K. J. Haworth, K. P. Mercado-Shekhara, and D. R. Dowling, “Frequency-sum beamforming for passive cavitation imaging,” *J. Acoust. Soc. Am.*, vol. 144, no. 1, p. 198, Jul 2018.
  - [51] J. Baranger, B. Arnal, F. Perren, O. Baud, M. Tanter, and C. Demen, “Adaptive spatiotemporal SVD clutter filtering for ultrafast doppler imaging using similarity of spatial singular vectors,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1574–1586, 2018.
  - [52] M. D. Gray, E. Lyka, and C. C. Coussios, “Diffraction effects and compensation in passive acoustic mapping,” *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 65, no. 2, pp. 258–268, 2018.
  - [53] 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 transactions on medical imaging*, vol. 37, no. 12, pp. 2582–2592, 2018.
  - [54] 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.*, 2018.
  - [55] Z. Wang, J. Li, P. Stoica, T. Nishida, and M. Sheplak, “Constant-beamwidth and constant-powerwidth wideband robust Capon beamformers for acoustic imaging,” *J. Acoust. Soc. Am.*, vol. 116, no. 3, p. 1621, Sep 2004.
  - [56] F. J. Harris, “On the use of windows for harmonic analysis with the discrete Fourier transform,” *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, Jan 1978.
  - [57] L. Petrusca, F. Varray, R. Souchon, A. Bernard, J.-Y. Chapelon, H. Liebgott, W. A. NDjin, and M. Viallon, “Fast volumetric ultrasound b-mode and doppler imaging with a new high-channels density platform for advanced 4d cardiac imaging/therapy,” *Applied Sciences*, vol. 8, no. 2, 2018.