# Neural Optimisation of Fixed Beamformers With Flexible Geometric Constraints

Longfei Felix Yan, *Member, IEEE*, Weilong Huang, Thushara D. Abhayapala, *Fellow, IEEE*, Jinwei Feng, and W. Bastiaan Kleijn, *Fellow, IEEE*

***Abstract*—This paper presents a novel approach to optimising fixed broadband beamformers using neural networks. The proposed neural network model for fixed beamformers allows for the optimisation of spatial filters while incorporating flexible geometric constraints. We propose a framework for the unified signal model applicable to all geometric settings and employ two heterogeneous neural networks to simultaneously optimise both the geometry and spatial filter of fixed beamforming. Furthermore, we introduce a technique called constrained naked neurons for the optimisation of spatial filters. Experimental results show that our approaches outperform conventional approaches in terms of Directivity Factor (DF) and White Noise Gain (WNG). Our study reveals the competitive performance of a circular microphone array that matches the capabilities of a concentric circular microphone array with the same number of microphones. We also validate the effectiveness of our model in a circular discal setting, where microphones can be placed arbitrarily. Given the same parameter settings, a circular discal array can be significantly better than a linear array.**

***Index Terms*—Directivity factor, fixed beamforming, frequency-invariant beampatten, geometry optimisation, microphone arrays, neural network, white noise gain.**

## I. Introduction

MICROPHONE arrays are indispensable in various hands-free communication systems and far-field speech recognition systems, especially in challenging environments [1], [2], [3], [4], [5]. Devices equipped with microphone arrays hold potential for applications such as speaker separation [6], [7], speaker localization [8], [9], speech dereverberation [10], and speech enhancement [11], [12]. Beamforming, as a critical function of microphone arrays, has garnered significant attention [13], [14]. It acts as a spatial filtering process, enhancing the target signal from the desired direction while suppressing interference from undesired directions [15].

Received 27 February 2024; revised 28 October 2024; accepted 9 January 2025. Date of publication 24 January 2025; date of current version 12 February 2025. The associate editor coordinating the review of this article and approving it for publication was Dr. Romain Serizel. *(Corresponding author: Weilong Huang.)*

Longfei Felix Yan and W. Bastiaan Kleijn are with the School of Engineering and Computer Science, Victoria University of Wellington, Wellington 6140, New Zealand.

Weilong Huang is with the International Audio Laboratories Erlangen (a Joint Institution of Friedrich-Alexander-Universität Erlangen-Nürnberg and Fraunhofer IIS), 91058 Erlangen, Germany (e-mail: weilong.huang89@gmail.com).

Thushara D. Abhayapala is with the School of Engineering, the Australian National University, Canberra, ACT 2601, Australia.

Jinwei Feng is with Alibaba Group, Bellevue, WA 98004 USA.

Digital Object Identifier 10.1109/TASLPRO.2025.3533372

Differential Microphone Arrays (DMAs) are motivated by the spatial derivative of the acoustic pressure field [16]. In the realm of fixed beamforming, there has been a growing interest in the application of DMAs due to their distinct advantages over certain traditional beamformers [13], [17], [18], [19]. First, DMAs can generate a relatively frequency-invariant beampattern, making them well-suited for broadband speech processing. Second, DMAs have the potential to achieve high Directivity Factors (DFs) with their compact apertures. Lastly, by incorporating appropriate optimisation constraints [17] or predefined target beampatterns [20], DMAs can achieve a more balanced combination of DFs and White Noise Gains (WNGs) compared to other fixed beamformers like Delay-and-Sum and superdirective beamformers [21]. However, the predefined target beampatterns are not guaranteed to be optimal. The beampattern parameters such as null directions are often manually selected by human experts.

Various DMA geometries have been studied, including uniform linear arrays [18], [22], nonuniform linear arrays [23], [24], uniform circular arrays [25], and uniform concentric circular arrays [26]. When it comes to a microphone array with a fixed geometry, designing DMAs typically involves striking a balance between WNGs and DFs [14]. Robust DMAs, known to achieve maximum WNG with a minimum norm solution, tend to have a lower DF. On the other hand, Maximum Directivity (MDF) differential beamformers can attain a high DF at the expense of WNG. To address this trade-off, a parameterized DMA has been proposed in [27], where DF and WNG are compromised based on a parameter. The order of DMAs also affects the WNG and DF performance and it is expressed by the order of the MacLaurin series used to approximate the exponential term [21], [28]. Each order of DMAs requires a minimum number of microphones to achieve [21]. It is possible to improve WNG without sacrificing DF by employing more microphone elements than what is required for the order of robust DMAs [20], [29] or by leveraging the acoustic properties of directional sensors [30], [31], [32], [33]. Similarly, DF can be enhanced without sacrificing WNG through the use of acoustic vector sensors [34], [35]. However, such improvements come at a financial cost in practice. Incorporating more microphones not only increases the cost of the microphones themselves, but also leads to a more complex hardware architecture. Furthermore, the implementation of directional sensors in products necessitates sound transparency, resulting in a much more intricate industrial design compared to omnidirectional sensors.

As mentioned above, the majority of recent DMAs have been developed on the basis of specific given geometries, without considering the optimisation of geometry itself. Optimisation of geometry design for DMA beamformers has been explored mainly for linear microphone arrays using techniques such as particle swarm optimisation (PSO) [36], [37] or convex optimisation [38], [39]. In the PSO-based approach proposed in [36], DMA beamformers demonstrated a superior tradeoff between DF and WNG compared to traditional methods. However, the distortionless constraint in the desired direction is not guaranteed, leading to power distortion across different frequency bins, deteriorating the quality of perceived broadband signals. Additionally, these PSO-based DMA beamformers suffer from white noise amplification in low-frequency bins. To address these limitations, a neural optimisation method was proposed for fixed beamformers on linear arrays utilising machine learning techniques [40]. This approach shows an overall improvement in WNG and DF compared to PSO-based DMA beamformers. However, this method is limited to linear arrays and does not provide separate improvements for WNG and DF individually.

Building on the work presented in [40], we propose a novel neural network-based method that enables joint optimisation of the array geometry and fixed beamformers. This method caters to a range of geometries, including linear, circular, and concentric circular arrays. The key innovation lies in integrating the ResNet structure [41] and the augmented Lagrangian approach within the loss function to facilitate the optimisation process. ResNets are well known for their proven global convergence property [42], [43]. Inspired by this, we reformulate the constrained optimisation problem associated with the performance of fixed beamformers within a neural network framework.

Our framework focuses on enhancing the effectiveness of the optimisation process by leveraging the power of the nonlinear approximation of neural networks [44]. Neural networks construct solutions that optimise geometric states and spatial filters, enabling us to achieve superior performance in the beamforming process. To validate the effectiveness of our neural network framework, a comparative analysis with Constrained Naked Neurons (ConNNs) is performed. In ConNNs, only stochastic gradient descent is employed within predefined constraints to optimise fixed beamformers.

Our contributions can be summarised as follows: (i) We introduce the Geometrically Optimised Neural Fixed Beamformer (GONFB), which optimises the beampatterns of fixed beamformers and array geometries using stochastic gradient descent. (ii) We reveal the competitiveness of a circular microphone array. (iii) We eliminate deep nulls that are commonly encountered in DMA beamformers. (iv) We demonstrate that circular discal arrays can outperform linear arrays in a fair comparison. (v) We compare the fixed beamformers designed by GONFB, DMA, and ConNNs on different types of arrays. Experimental results show that GONFB exhibits significant improvements in both WNG and DF compared to current DMA-based methods. While ConNNs exhibit commendable proficiency in optimising spatial filters, they fall short when optimising geometry, a domain in which our proposed method, GONFB, excels.

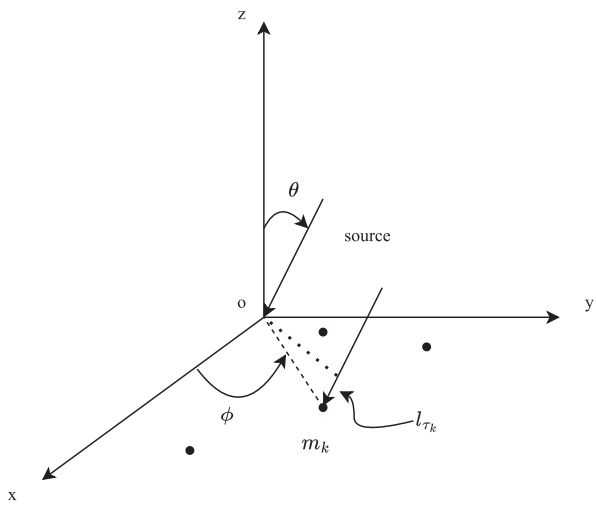



Fig. 1. The unified coordinate system in a 3D space. Black dots represent microphones. $\theta$ denotes the elevation angle and $\phi$ denotes the azimuth angle. $l_{\tau_k}$ represents the scalar projection of the microphone location vector $\mathbf{p}_k$ over the plane wave direction $\hat{\mathbf{a}}$.

The remaining sections of this paper are organised as follows: Section II presents a detailed description of our signal model and problem formulation. Section III contextualises related DMA beamformers within our research framework. In Section IV, we outline the performance measures utilised in this study. Building upon this, Section V introduces our neural network model, which optimises both array geometries and spatial filters of fixed beamformers. Section VI provides a comprehensive discussion on our experimental settings and the corresponding results. Finally, in Section VII, we draw meaningful conclusions based on our findings.

## II. Signal Model and Problem Formulation

The objective of beamforming is to enhance signals in the target direction while attenuating signals from other directions. The geometry of a microphone array plays a fundamental role in beamforming. Practically, the microphone array for a fixed beamformer can form any geometry in a 3D space. To accommodate the flexible geometry of real-world microphone arrays, we propose to unify the coordinate system for fixed beamformers and consider an array of $M$ microphones with an arbitrary geometry, as shown in Fig. 1. Let the $k$th microphone, denoted by $m_k$, be located at $\mathbf{p}_k = [x_k, y_k, z_k]^T$. The direction of a plane wave can be represented by a unit vector

$$\hat{\mathbf{a}} = [-\sin\theta\cos\phi \;\; -\sin\theta\sin\phi \;\; -\cos\theta]^T, \tag{1}$$

where $\theta$ is the elevation angle and $\phi$ is the azimuth angle of the plane wave. Using the geometric definition of the dot product, we can derive the scalar projection of $\mathbf{p}_k$ over $\hat{\mathbf{a}}$, denoted by $l_{\tau_k}$ as:

$$\begin{aligned} l_{\tau_k} &= \mathbf{p}_k^T \hat{\mathbf{a}} \\ &= -(x_k \sin\theta\cos\phi + y_k \sin\theta\sin\phi + z_k \cos\theta). \end{aligned} \tag{2}$$

Let $c$ be the speed of sound in the air. The difference in arrival time between the origin $o$ and the microphone $m_k$ is $\tau_k = l_{\tau_k}/c$. Thus, the steering vector [15], [45] can be defined as

$$\begin{aligned}\mathbf{d}(\omega, \hat{\mathbf{a}}) &= [e^{-\jmath\omega \mathbf{p}_1^T \cdot \hat{\mathbf{a}}/c} \ \cdots \ e^{-\jmath\omega \mathbf{p}_M^T \cdot \hat{\mathbf{a}}/c}]^T \\ &= [e^{-\jmath\omega\tau_1} \ \cdots \ e^{-\jmath\omega\tau_M}]^T, \end{aligned} \tag{3}$$

where $\jmath = \sqrt{-1}$, $\omega = 2\pi f$ and $f$ is the temporal frequency.

To formulate the beamforming problem, we first define $s(\omega)$ as the far-field source signal. Under free-field propagation conditions, the received signal vector $\mathbf{y}(\omega)$ is defined as

$$\begin{aligned}\mathbf{y}(\omega) &= [y_1(\omega) \ y_2(\omega) \ \cdots \ y_M(\omega)]^T \\ &= \mathbf{d}(\omega, \hat{\mathbf{a}}) s(\omega) + \mathbf{v}(\omega), \end{aligned} \tag{4}$$

where $y_k(\omega)$ is the signal received at the $k$th microphone, and $\mathbf{v}(\omega)$ is the noise vector. By applying a beamforming filter $\mathbf{h}(\omega)$ to $\mathbf{y}(\omega)$, $\hat{s}(\omega)$ can be estimated:

$$\begin{aligned}\hat{s}(\omega) &= \mathbf{h}^H(\omega)\mathbf{y}(\omega) \\ &= \mathbf{h}^H(\omega)\mathbf{d}(\omega, \hat{\mathbf{a}}) s(\omega) + \mathbf{h}^H(\omega)\mathbf{v}(\omega). \end{aligned} \tag{5}$$

In this way, a beamformer tries to recover $s(\omega)$ from estimating $\hat{s}(\omega)$.

The filter $\mathbf{h}(\omega)$ in fixed beamforming is a vector of complex weights that performs spatial filtering at angular frequency $\omega$. The constraint of fixed beamformer design is to obtain the filter $\mathbf{h}(\omega)$ that achieves a unity gain at a desired look direction, $\hat{\mathbf{a}_\ell}$, over the interested frequency bands. In other words, the filter $\mathbf{h}(\omega)$ should satisfy the distortionless constraint at the desired look direction $\hat{\mathbf{a}_\ell}$, which is

$$\mathbf{h}^H(\omega)\mathbf{d}(\omega, \hat{\mathbf{a}_\ell}) = 1 \quad \forall\omega. \tag{6}$$

The value of $\mathbf{h}^H(\omega)\mathbf{d}(\omega, \hat{\mathbf{a}})$ should be less than 1 if $\hat{\mathbf{a}} \neq \hat{\mathbf{a}_\ell}$.

### A. Performance Measures

The array response, or beampattern, characterises the response of a microphone array as a function of the direction of incident sound waves [46]. It yields a graphical representation of the spatial sensitivity of the array to signals arriving from different directions. The beampattern for a plane wave of angular frequency $\omega$ arriving from a direction $\hat{\mathbf{a}}$ is given by

$$\mathcal{B}[\mathbf{h}(\omega), \hat{\mathbf{a}}] = \mathbf{h}^H(\omega)\mathbf{d}(\omega, \hat{\mathbf{a}}). \tag{7}$$

The robustness of a microphone array is indicated by White Noise Gain (WNG) [47]. The higher the WNG is, the more resilient the microphone arrays are against self-noise and microphone mismatches in the desired look direction. The formula of WNG is

$$\mathrm{WNG}[\mathbf{h}(\omega), \hat{\mathbf{a}}_\ell] = \frac{|\mathcal{B}[\mathbf{h}(\omega), \hat{\mathbf{a}}_\ell]|^2}{\mathbf{h}^H(\omega)\mathbf{h}(\omega)}. \tag{8}$$

The directionality of a microphone array is measured by Directivity Factor (DF). It is a ratio of the output power in the direction of interest to the total output power. A high DF suggests that the array can focus on a particular direction while suppressing energy from other directions. It is defined as

$$\mathrm{DF}[\mathbf{h}(\omega), \hat{\mathbf{a}}_\ell] = \frac{|\mathcal{B}[\mathbf{h}(\omega), \hat{\mathbf{a}}_\ell]|^2}{\mathbf{h}^H(\omega)\mathbf{\Gamma}_{0,\pi}(\omega)\mathbf{h}(\omega)}, \tag{9}$$

where $\mathbf{\Gamma}_{0,\pi}(\omega)$ is a square matrix of size $M$. The elements in $\mathbf{\Gamma}_{0,\pi}(\omega)$ are given by

$$[\mathbf{\Gamma}_{0,\pi}(\omega)]_{ij} = \mathrm{sinc}[\omega||\mathbf{p}_i - \mathbf{p}_j||_2/c], \tag{10}$$

where $\mathrm{sinc}(x) = \sin x/x$.

### B. Special Cases

The definition of the steering vector in (3) is a general description of the spatial information of a microphone array in any geometry. Here, we provide the steering vectors for frequently used array geometries.

*1) Linear Array:* Without loss of generality, we can align the Linear Array (LA) with the $z$-axis. Its location can be expressed as $\mathbf{p}_{\mathrm{LA},k} = [0, 0, -q_k]^T$. This leads to the commonly used steering vector formula for linear arrays [17], [37]:

$$\mathbf{d}_{\mathrm{LA}}(\omega, \theta) = [1 \ e^{-\jmath\omega q_2 \cos\theta/c} \ \cdots \ e^{-\jmath\omega q_M \cos\theta/c}]^T. \tag{11}$$

*2) Circular Array:* Without loss of generality, we consider a Uniform Circular Microphone Array (UCMA) of radius $r$ with $M$ microphones in the $x$-$y$ plane. The centre of the circular array coincides with the coordinate origin. The $k$th microphone has a location $\mathbf{p}_{\mathrm{CA},k} = r[\cos\psi_k, \sin\psi_k, 0]^T$. The steering vector of a circular array can be derived as [48]:

$$\mathbf{d}_{\mathrm{CA}}(\omega, \phi) = [e^{\jmath\omega\cos(\phi-\psi_1)r/c} \ \cdots \ e^{\jmath\omega\cos(\phi-\psi_M)r/c}]^T, \tag{12}$$

where $\psi_k$ is the angular position of the $k$th microphone measured anticlockwise from the $y$ axis.

*3) Concentric Circular Array:* In the case of a Uniform Concentric Circular Microphone Array (UCCMA) of radii $r_i$, $i = 1, \ldots, I$, the location of the $k$th microphone on the $i$th ring is $\mathbf{p}_{\mathrm{CCA},i,k} = r_i[\cos\psi_{i,k}, \sin\psi_{i,k}, 0]^T$ in the $x$-$y$ plane, where $\psi_{i,k}$ is the angular position of the $k$th microphone on the $i$th ring. We derive the steering vector of the $i$th ring as:

$$\mathbf{d}_{\mathrm{CCA},i}(\omega) = [e^{\jmath\omega\cos(\phi-\psi_{i,1})r_i/c} \ \cdots \ e^{\jmath\omega\cos(\phi-\psi_{i,M_i})r_i/c}]^T, \tag{13}$$

where $M_i$ is the total number of microphones on the $i$th ring. The steering vector of the whole UCCMA is:

$$\mathbf{d}_{\mathrm{CCA}}(\omega) = [\mathbf{d}_{\mathrm{CCA},1}^T(\omega), \ldots, \mathbf{d}_{\mathrm{CCA},I}^T(\omega)]. \tag{14}$$

*4) Circular Discal Array:* As to Circular Discal Arrays (CDAs), consider the case of $M$ microphones located at arbitrary points $\mathbf{p}_{\mathrm{CDA},k} = r_k[\cos\psi_k, \sin\psi_k, 0]^T$ on a circular disk in the $x$-$y$ plane. The centre of the circular disk is at the coordinate origin. Its steering vector is:

$$\mathbf{d}_{\mathrm{CDA}}(\omega) = [e^{\jmath\omega\cos(\phi-\psi_1)r_1/c} \ \cdots \ e^{\jmath\omega\cos(\phi-\psi_M)r_M/c}]^T. \tag{15}$$

## III. Related Works

DMAs are popular fixed beamformers with desirable properties like relative frequency-invariant beampatterns [14], [17], [20], [21], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32],

[37], [49]. Based on (7), the beampattern of an $N$th order uniform linear DMA in the $x$-$y$ plane can be approximated as [17], [23], [37]:

$$\mathcal{B}_N(\phi) \approx \sum_{n=0}^{N} a_{N,n} \cos^n \phi, \tag{16}$$

where $N$ is the order of DMA and $a_{N,n}$ is the $n$th coefficient of the beampattern. This expression is derived by approximating exponential terms in the steering vector with MacLaurin series:

$$e^{\jmath\omega q_i \cos\phi/c} \approx \sum_{n=0}^{N} \frac{1}{n!}[\jmath\omega q_i \cos\phi/c]^n. \tag{17}$$

Note that the order of DMA is the same as the order of MacLaurin series used.

Some "ideal" beampatterns have been summarised for DMAs of different orders, such as hypercardioid and supercardioid. They are determined by the values of null directions, where zeros are placed at those directions. However, an "ideal" beampattern is optimal only with respect to one performance measure. For example, a hypercardioid maximizes DF but has poor WNG [21]. Furthermore, the choice for the order of DMA, corresponding to the number of null directions, is another issue. A higher order DMA is better at suppressing interfering noise, but less robust [21].

One popular technique for expressing exponential terms in DMA beampatterns is to use the Jacobi-Anger expansion [20]. In a uniform circular DMA, we have:

$$e^{\jmath\overline{\omega}\cos(\phi-\psi_k)} = J_0(\overline{\omega}) + 2\sum_{n=1}^{\infty} \jmath^n J_n(\overline{\omega}) \cos[n(\phi-\psi_k)], \tag{18}$$

where $\overline{\omega} = \omega r/c$ and $J_n(\overline{\omega})$ is the $n$-th order Bessel function of the first kind. The approximated beampatterns replace infinity in (18) with the order $N$ and exhibit least-square errors in relation to the "ideal" beampatterns. For the approximation, a $(2N+1) \times (2N+1)$ diagonal matrix composed of $J_n(\overline{\omega})$ is introduced:

$$\mathbf{J}(\overline{\omega}) = \text{diag}\left[\frac{1}{\jmath^{-N}J_{-N}(\overline{\omega})}, \ldots, \frac{1}{J_0(\overline{\omega})}, \ldots, \frac{1}{\jmath^N J_N}(\overline{\omega})\right]. \tag{19}$$

In this diagonal matrix, the occurrence of zeros of the Bessel function gives rise to the issue of deep nulls [48], where the zero denominators causes numerical instability in $\mathbf{J}(\overline{\omega})$. From the perspective of least-squares approximation error, we compute the minimum-norm filter $\mathbf{h}_{\text{JA}}(\omega)$ with the Jacobi-Anger expansion [20]. A few more terms related to the azimuth angle are defined. The first one is:

$$\Upsilon(\psi_\ell) = \text{diag}(e^{\jmath N\psi_\ell}, \ldots, 1, \ldots, e^{-\jmath N\psi_\ell}), \tag{20}$$

where $\psi_\ell$ is the look direction of the azimuth angle. The second one is:

$$\Psi = [\boldsymbol{\psi}_{-N}^H, \ldots, \boldsymbol{\psi}_0^H, \ldots, \boldsymbol{\psi}_N^H]^T, \tag{21}$$

where $\boldsymbol{\psi}_n = [e^{\jmath n\psi_1}\ e^{\jmath n\psi_2}\ \cdots\ e^{\jmath n\psi_M}]$. Finally, we can derive the minimum-norm filter $\mathbf{h}_{\text{JA}}(\omega)$:

$$\mathbf{h}_{\text{JA}}(\omega) = \Psi^H(\Psi\Psi^H)^{-1}\mathbf{J}^H(\overline{\omega})\Upsilon^H(\psi_\ell)\mathbf{b}_{2N}, \tag{22}$$

where $b_{2N,i} = b_{2N,-i} = \frac{1}{2}a_{N,i}$ for $i = 0, 1, 2, \ldots, N$.

As mentioned above, deep nulls are numerical explosion phenomena caused by zero values of Bessel functions, which are inevitable in the Jacobi-Anger expansion-based DMA formulation. Specifically, the zero values of Bessel functions would make values of $\mathbf{J}^H(\overline{\omega})$ in (22) explosively large. Consequently, the norm of $\mathbf{h}_{\text{JA}}(\omega)$ becomes too large and the WNG value in (8) becomes too small. Deep nulls significantly undermine the performance of the beamformer at specific frequencies. To mitigate deep nulls, an alternative way to design DMA beamformers is by using the null-constrained approach [17]. Inspired by the observation that an $N$th-order DMA can have $N$ distinct null directions, the null-constrained approach selects $N$ null directions manually. In the $x$-$y$ plane, a matrix $\mathbf{D}(\omega)$ can be constructed as:

$$\mathbf{D}(\omega) = \begin{bmatrix} \mathbf{d}^H(\omega, \phi_\ell) \\ \mathbf{d}^H(\omega, \phi_1) \\ \vdots \\ \mathbf{d}^H(\omega, \phi_N) \end{bmatrix}, \tag{23}$$

where $\phi_\ell$ is the look direction and $\phi_1, \cdots, \phi_N$ are $N$ distinct null directions. Thus, $\mathbf{h}(\omega)$ can be derived by solving

$$\mathbf{D}(\omega)\mathbf{h}(\omega) = \mathbf{i}, \tag{24}$$

where $\mathbf{i} = [1\ 0\ \cdots\ 0]^T$ is a one-hot vector of length $N+1$. The minimum-norm solution to (24) in the null-constrained approach is:

$$\mathbf{h}_{\text{NC}}(\omega) = \mathbf{D}^H(\omega)[\mathbf{D}(\omega)\mathbf{D}^H(\omega)]^{-1}\mathbf{i}. \tag{25}$$

From (25), we can see that it is crucial to decide the number and value of null directions in order to derive $\mathbf{h}_{\text{NC}}(\omega)$. Yet there is no good mechanism of selecting parameters for null directions other than manually selecting some.

A recent study by the authors in [48] has illuminated that a null-constrained DMA can be viewed as a regularised Jacobi-Anger expansion-based DMA. In a unified framework, (25) can be approximated by Bessel functions with additional regularisation terms. The deep nulls arising from the zero values of Bessel functions in the Jacobi-Anger expansion-based DMA are mitigated through the presence of regularisation term in the null-constrained DMA. Nevertheless, it is imperative to acknowledge that the effectiveness of regularisation terms are dependent upon factors such as the number of microphones and the order of DMA. For example, when the number of microphones $M$ is odd and the order is $(M-1)/2$, null-constrained DMA also suffers from deep nulls [48].

In summary, the limitations of DMA beamformers in selecting optimal parameters, such as the order of DMA and null directions, impede their ability to provide optimal solutions for beamforming tasks. Moreover, deep nulls persist in all variants of DMA approaches, and they can only be mitigated in certain

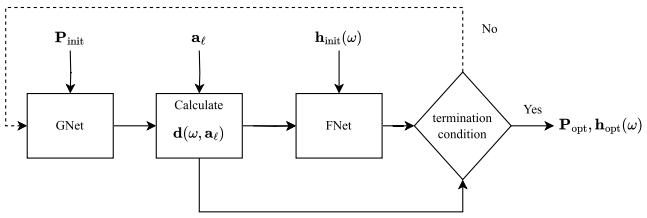



Fig. 2. Workflow diagram of GONFB, which consists of iterative optimisation of both GNet and FNet. The dashed line indicates that no actual data is transmitted.

cases. In contrast, machine learning models, such as neural networks, excel at optimising parameters and searching for near-optimal solutions when a globally optimal solution cannot be achieved analytically. Neural networks do not rely on Bessel functions have the potential to eliminate deep nulls. As a result, there is a need for a neural network model that can enhance the performance of fixed beamformers.

## IV. Neural Network Model and Constrained Naked Neurons

In this section, we introduce the architecture of the neural network model and the loss function formulated from the augmented Lagrangian. We also explain the motivation, definition, and implementation of constrained naked neurons as one of our baseline methods.

### A. Neural Network Model Architecture

Deep neural networks are known for their universal approximation when combined with nonlinear activation functions [50]. Previous research efforts have demonstrated the excellent convergence characteristics of neural network architectures such as ResNet in the pursuit of global optima [42], [43]. Consequently, employing a ResNet-like neural network model for the optimisation of fixed beamformers should be no exception when seeking near-optimal solutions. To show that the efficacy of our neural network model is more than stochastic gradient descent, we introduce ConNNs for comparative analysis, as elucidated in Section IV-B.

Inspired by the success of statistics networks [40], [51], [52], we propose a neural network model called Geometrically optimised Neural Fixed Beamformer, abbreviated as GONFB. The full model consists of two heterogeneous neural networks, GNet and FNet. GNet is a fully connected feed-forward neural network, and FNet has an architecture resembling ResNet. The model is trained in a cascaded manner, as shown in Fig. 2. GNet takes initial geometry inputs, that is, the initial coordinates of the microphones $\mathbf{P}_{\text{init}} = [\mathbf{p}_1, \ldots, \mathbf{p}_M]^T$ of size $M \times 3$, and produces optimised microphone coordinates $\mathbf{P}_{\text{opt}}$. The steering vector $\mathbf{d}(\omega, \hat{\mathbf{a}}_\ell)$ can be calculated by utilising $\mathbf{P}_{\text{opt}}$ and the direction of the signals $\hat{\mathbf{a}}_\ell$. Subsequently, FNet takes $\mathbf{d}(\omega, \hat{\mathbf{a}}_\ell)$ and the initial filter $\mathbf{h}_{\text{init}}(\omega)$ to produce the optimised filter $\mathbf{h}_{\text{opt}}(\omega)$. During training, early termination is applied if the loss decreases in the next evaluation. Otherwise, training will continue until the maximum epoch number is reached. During each iteration, the weights of GNet and FNet are updated by gradient descent in an unsupervised manner. The inputs $\mathbf{P}_{\text{init}}$, $\mathbf{h}_{\text{init}}(\omega)$, and $\hat{\mathbf{a}}_\ell$ remain the same to facilitate neural network training.

*1) GNet Implementation Details:* GNet has three layers in total: an input linear layer, a hidden layer and an output linear layer. A linear layer multiplies the input with a weight matrix. The input linear layer allows for the transformation of the input into a new latent representation that captures important features of the data. The output linear layer extracts the latent representation with the desired dimensions. In our experiments, it turns out that the optimisation of geometry requires significantly fewer parameters than the optimisation of array filters. This is because 2D array geometry has limited complexity in an x-y plane. The nonlinear activation functions in the input linear layer and the hidden layer are Rectified Linear Units (ReLU) [53]. They are widely used in neural networks because of their simplicity and effectiveness. A ReLU function is expressed as:

$$\text{ReLU}(u) = \max(0, u), \tag{26}$$

where positive inputs are intact and negative inputs are converted to 0.

For a linear array, a softmax function is applied, defined as:

$$\sigma(\mathbf{u})_i = \frac{e^{\mathbf{u}_i}}{\sum_{j=1}^{M-1} e^{\mathbf{u}_j}}, \tag{27}$$

where $\mathbf{u}$ is the input latent embedding vector and $\mathbf{u}_i$ is its $i$th element. The softmax function has the desirable property that the sum of its elements is equal to 1, similar to proportions. This property enables us to assign spacing between microphones based on the proportion of each element in the output vector. In other words, each element's value in the softmax function represents its relative position along the length of the linear array.

A sigmoid function, defined as

$$\text{SIG}(u) = \frac{1}{1 + e^{-u}}, \tag{28}$$

is used for a UCCMA instead. As the sigmoid function's range is between 0 and 1, it can represent the proportion of an attribute by using its output, provided that the maximum value of the attribute is known. For instance, in a UCCMA, the angular position of the first microphone on a ring is at most $2\pi/M$, where $M$ microphones are uniformly spaced. The angular position of the first microphone corresponds to a proportion of $2\pi/M$. Optimisation of the radius of rings is another geometry parameter in a UCCMA. The proportion of the maximum radius for different rings is optimised, given that the maximum length of the radius is known.

*2) FNet Implementation Details:* FNet comprises an input linear layer, an output linear layer, and multiple ResBlocks in between, as shown in Fig. 3. All layers in FNet handle complex data by enabling independent processing of the real and imaginary parts of the input [54]. Specifically, imaginary parts of the input are treated as though they were real. A ResBlock is a building block inspired from ResNet [41], as shown in Fig. 4. We employ the residual learning mechanism in ResBlocks, which adds an identity mapping between input and output, to ease the training of deep neural networks. The Gaussian Error Linear

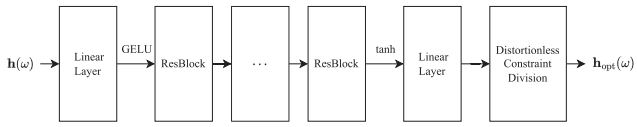



Fig. 3. The diagram of FNet, which comprises multiple ResBlocks between the input and output linear layers. The distortionless constraint is enforced in the end.

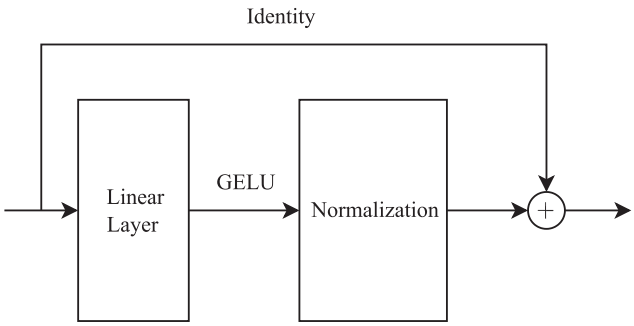



Fig. 4. The diagram of a ResBlock. An identity mapping is added after linear transformation, GELU and normalisation.

Unit (GELU) function is expressed as [55]

$$\text{GELU}(u) = u\Phi(u) = u \cdot \frac{1}{2}[1 + \text{erf}(u/\sqrt{2})], \tag{29}$$

where $u$ is the input of the function, $\Phi(u)$ is the standard Gaussian cumulative distribution function and it can be algebraically represented by a manipulation of the error function

$$\text{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-t^2} dt. \tag{30}$$

The GELU function scales inputs by their value instead of gating inputs by their signs. This has shown superior performance than the ReLU and ELU functions in computer vision, natural language processing, and speech tasks [55], [56], [57], [58]. The hyperbolic tangent function (tanh) is a commonly used activation function in neural networks. Its range is between $-1$ and 1 and helps to prevent the saturation of neurons. The normalisation layer is incorporated to facilitate the optimisation process by centering the input data and scaling it to have unit variance [59]. Specifically, the normalisation layer transforms the input by subtracting its mean and dividing it by its standard deviation, resulting in a zero mean and unit variance representation. Finally, to enforce the distortionless constraint, we divide $\mathbf{h}_{\text{opt}}(\omega)$ by the product $\mathbf{h}_{\text{opt}}^H(\omega)\mathbf{d}(\omega, \hat{\mathbf{a}}_\ell)$.

### B. Constrained Naked Neurons

In this work, we introduce a structure named Constrained Naked Neurons (ConNNs), which diverges from the conventional deep learning paradigm. The term "naked" is used to convey that these neurons lack both learnable weights and bias terms. Additionally, these naked neurons are devoid of interconnections to other neurons; their role is solely to impose constraints on the inputs using fixed functions. For instance, activation functions such as sigmoid in (28) or ReLU in (26) can be applied to ConNNs. Instead of iteratively adapting weights and bias terms, ConNNs focus on the direct optimisation of inputs. By subjecting initial inputs to feasible constraints and employing the Stochastic Gradient Descent (SGD) algorithm, ConNNs iteratively refine inputs through backpropogation to align with a predefined loss function. Consequently, the optimisation process hinges on the initial input data, the applied fixed function, and the predefined loss function. This framework is especially suitable for constrained optimisation problems where access to extensive training data is limited. As ConNNs are also driven by SGD, they serve as valuable baselines for comparison against neural networks. Such a comparison elucidates the distinct contributions of neural networks on top of the constraints, the loss function and the SGD algorithm.

Similar to the neural network architecture, the input elements of ConNNs encompass the initial filter denoted as $\mathbf{h}_{\text{init}}(\omega)$ and the initial coordinates of the microphones denoted as $\mathbf{P}_{\text{init}}$. The filter's constraint aligns with the distortionless criterion articulated in (6), realized through the division of $\mathbf{h}_{\text{opt}}(\omega)$ by the product $\mathbf{h}_{\text{opt}}^H(\omega)\mathbf{d}(\omega, \hat{\mathbf{a}}_\ell)$. Geometric constraints vary across distinct spatial configurations. For linear arrays, $\mathbf{P}_{\text{init}}$ adheres to the array's length constraint, where the summed spacing between microphones derived from $\mathbf{P}_{\text{opt}}$ corresponds to the designated array length $A$. A ratio $\rho = A/\widehat{A}$ is computed, with $\widehat{A}$ representing the estimated array length obtained from ConNNs. This ratio is multiplied to the estimated microphone spacing to adjust the final outputs. In the context of circular arrays, the array geometry is predefined as a uniform circular array, obviating the need for geometric optimisation. In concentric circular arrays, a sigmoid function defined in (28) constrains the starting microphone angular position and the radii of distinct rings. This sigmoid function yields proportional parameter values relative to the maximum values.

The implementation of ConNNs is straightforward, necessitating only initial inputs, constraints, and a loss function. The SGD algorithm drives ConNNs towards improved solutions from the initial inputs. However, it's worth noting that SGD can become ensnared in local optima. Consequently, the final output of ConNNs is dependent on the quality of the initial inputs provided.

### C. Loss Function

For convenience of presentation, we denote $\text{WNG}[\mathbf{h}(\omega)]$ as $f_i(\mathcal{X})$ and $\text{DF}[\mathbf{h}(\omega)]$ as $g_i(\mathcal{X})$, where $i$ is the frequency bin index corresponding to $\omega$, and $\mathcal{X} = \mathbf{h}(\omega)$. Assuming that the distortionless constraint is satisfied, we can formulate the optimisation problem of our frequency-invariant beamformer as

$$\begin{aligned} &\min_{\Theta} -\frac{1}{F}\sum_{i=1}^{F} f_i(\mathcal{X}) \\ &\text{s.t. } \alpha_i \le g_i(\mathcal{X}) \le \beta_i, \quad i = 1, \ldots, F, \end{aligned} \tag{31}$$

where $\Theta$ represents the parameters to be optimised, $F$ is the number of frequency bins, $\alpha_i$ and $\beta_i$ are two-sided constraints of $g_i(\mathcal{X})$ with $\alpha_i < \beta_i$.

The augmented Lagrangian method is a powerful numerical optimisation technique that combines the Lagrangian function and a quadratic penalty function [60], [61]. It incorporates the constraints and the objective into a single function, which is suitable to be a loss function of a neural network. It is smoother and less ill-conditioned than using the penalty function directly [61]. In our case, the augmented Lagrangian dynamically adjusts the weights between DFs and WNGs with a second-order regularisation term. The objective optimises WNGs and the constraints take care of the target values of DFs. We can write (31) in an equivalent form:

$$\min_{\Theta} -\frac{1}{F}\sum_{i=1}^{F} f_i(\mathcal{X}) \quad \text{s.t. } \alpha_i \le g_i(\mathcal{X}) - u_i \le \beta_i,\ \ u_i = 0,\ i = 1,\ldots,F, \tag{32}$$

Subsequently, (32) can be converted to $\mathcal{L}(\mathcal{X})$[62]

$$\min_{\Theta} \mathcal{L}(\mathcal{X}) = -\frac{1}{F}\sum_{i=1}^{F} f_i(\mathcal{X}) + \sum_{i=1}^{F} p_i\left(g_i(\mathcal{X}), \mu_k^i, c_k^i\right), \tag{33}$$

where $\mathcal{L}(\mathcal{X})$ is the loss function of our neural network framework to be minimised and $\Theta$ denotes its parameters to be optimised, $i = 1,\ldots,F$, $\mu_k^i$ is the $i$th Lagrangian multiplier in the $k$th iteration, $c_k^i$ is the $i$th penalty coefficient in the $k$th iteration and

$$p_i\left(g_i(\mathcal{X}), \mu_k^i, c_k^i\right) = \min_{\alpha_i \le g_i(\mathcal{X}) - u_i \le \beta_i} \left\{\mu_k^i u_i + \frac{1}{2} c_k^i |u_i|^2\right\}. \tag{34}$$

$c_k^i$ is a positive large number intended to squash the value of $|u_i|^2$. Analytically, we can obtain the optimal solution [62] for (34)

$$u_i = \begin{cases} g_i(\mathcal{X}) - \beta_i & \text{if } \mu_k^i + c_k^i[g_i(\mathcal{X}) - \beta_i] > 0, \\ g_i(\mathcal{X}) - \alpha_i & \text{if } \mu_k^i + c_k^i[g_i(\mathcal{X}) - \alpha_i] < 0, \\ -\mu_k^i / c_k^i & \text{otherwise.} \end{cases} \tag{35}$$

Thus, (34) now becomes

$$p_i\left(g_i(\mathcal{X}), \mu_k^i, c_k^i\right) = \begin{cases} \mu_k^i[g_i(\mathcal{X}) - \beta_i] + \frac{1}{2} c_k^i |g_i(\mathcal{X}) - \beta_i|^2 & \text{if } \mu_k^i + c_k^i[g_i(\mathcal{X}) - \beta_i] > 0, \\ \mu_k^i[g_i(\mathcal{X}) - \alpha_i] + \frac{1}{2} c_k^i |g_i(\mathcal{X}) - \alpha_i|^2 & \text{if } \mu_k^i + c_k^i[g_i(\mathcal{X}) - \alpha_i] < 0, \\ -(\mu_k^i)^2 / 2c_k^i & \text{otherwise.} \end{cases} \tag{36}$$

Using the first-order derivative, the update rule for $\mu_k^i$ in the next iteration is

$$\mu_{k+1}^i = \begin{cases} \mu_k^i + c_k^i[g_i(\mathcal{X}) - \beta_i] & \text{if } \mu_k^i + c_k^i[g_i(\mathcal{X}) - \beta_i] > 0, \\ \mu_k^i + c_k^i[g_i(\mathcal{X}) - \alpha_i] & \text{if } \mu_k^i + c_k^i[g_i(\mathcal{X}) - \alpha_i] < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{37}$$

To update the penalty coefficient $c_k^i$, we can multiply it by 2 whenever the constraint for $g_i(\mathcal{X})$ is violated. This yields:

$$c_{k+1}^i = \begin{cases} \min(\tilde{c}, 2c_k^i) & \text{if } g_i(\mathcal{X}) > \beta_i \text{ or } g_i(\mathcal{X}) < \alpha_i \\ c_k^i & \text{otherwise,} \end{cases} \tag{38}$$

where $\tilde{c}$ is the maximum value for $c_k^i$. In our neural network framework, the loss function $\mathcal{L}(\mathcal{X})$ is minimised through the stochastic gradient descent by adjusting $\Theta$.

## V. Experiments

We conducted experiments on microphone arrays with four different geometry scenarios: linear, circular, concentric circular, and circular discal. The circular discal geometry had microphones distributed on a 2D circular plate. In each scenario, we reported the results from our best run. For the first three scenarios, we compared our approaches with state-of-the-art frequency-invariant beamformers [20], [26], [37]. For the circular discal geometry, we demonstrated that our approach could obtain nearly global optimal solutions without any pre-specified geometry constraints. Additionally, we compared the performance of our circular discal array with the corresponding linear array [37]. ConNNs shared the same parameters with GONFB where possible.

### A. Neural Network Training Details

The Adam optimiser [63] was used in all geometry scenarios. The learning rate was 0.00001. The gradient norm was clipped and the maximum gradient norm was 5. The parameters of the augmented Lagrangian in the first iteration were $\mu_0^i = 0$ and $c_0^i = 5$. Geometry-dependent parameters such as the number of neurons in each layer are provided in each geometry scenario specifically.

### B. Linear Array

*1) Initialisation and Hyperparameters:* The linear array to be optimised had 16 microphones. The minimum spacing between microphones was set at 0.4 cm. The length of the array was 15 cm. The initial filter was obtained by applying a null-constrained method [17], with two nulls at $106^\circ$ and $153^\circ$. The initial geometry was a uniform linear array. FNet had 10 hidden layers. Each layer had 180 neurons. The loss was evaluated every 40,000 epochs. The parameters of the augmented Lagrangian were updated every 10,000 epochs. The maximum epoch number was 400,000. The target DF value was 8.24 dB. With a tolerance of 0.1 dB, we set $\alpha_i = 8.14$ and $\beta_i = 8.34$. The frequency range was from 300 Hz to 8000 Hz. The look direction of the array was endfire. The PSO DMA method was from [37].

*2) Performance Analysis:* The optimised geometry of the linear array is shown in Fig. 5. Three subarrays are formed to handle the broadband signal processing by GONFB. The phenomenon that dense microphones are in the middle of the array has been observed by another frequency-invariant linear array approach [64]. Additionally, our approach placed two smaller dense subarrays at both ends of the linear array. The same

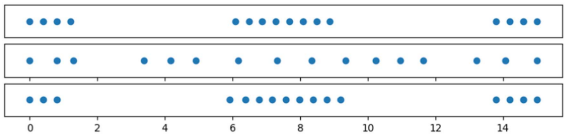


Fig. 5. optimised linear array geometry of GONFB (top), PSO DMA (middle), and ConNNs (bottom), when $M = 16$ and $L = 0.15$ m.

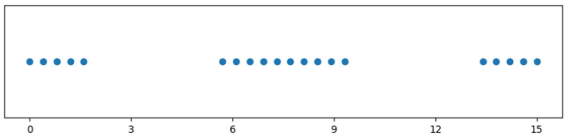


Fig. 6. optimised linear array geometry of GONFB when $M = 20$ and $L = 0.15$ m.

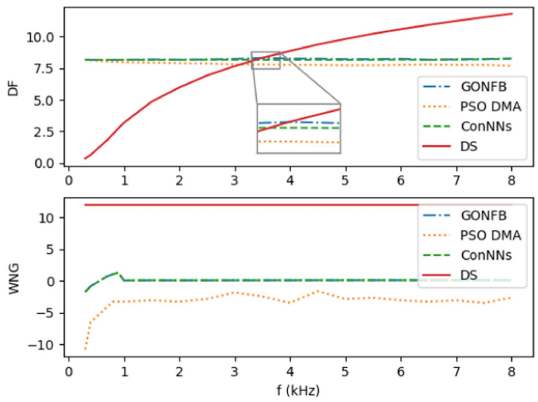



Fig. 7. Performance of GONFB, PSO DMA, ConNNs and DS in a linear array when $M = 16$ and $L = 0.15$ m.

geometric pattern is observed again when GONFB arranges 20 microphones in Fig. 6. The small spacing in each subarray helps process higher frequency signals and the large gap between subarrays can deal with lower-frequency signals. Similarly, three subarrays are also observed in the PSO DMA approach and the ConNNs approach. For the PSO DMA approach, the individual subarrays are not as dense. This could explain the inferior performance of DMAs as shown in Fig. 7, as more compact subarrays can better process high-frequency signals. For the ConNNs approach, the three subarrays are almost identical to the ones in GONFB except that there is one less microphone in the left subarray, which is shifted to the middle subarray. Apparently, ConNNs yield a local optimal solution and perform slightly worse than GONFB, which can be corroborated in Fig. 7.

We compare the performance of our linear array with the PSO-based frequency-invariant nonuniform linear DMA described in [37], the Delay-and-Sum (DS) approach [21], and ConNNs, as shown in Fig. 7. GONFB outperforms both the DMA-based approach and ConNNs in terms of DF and WNG values. While the DS approach shows good WNG performance, its DF is significantly lower in low-frequency bins and lacks frequency invariance. These results demonstrate that GONFB

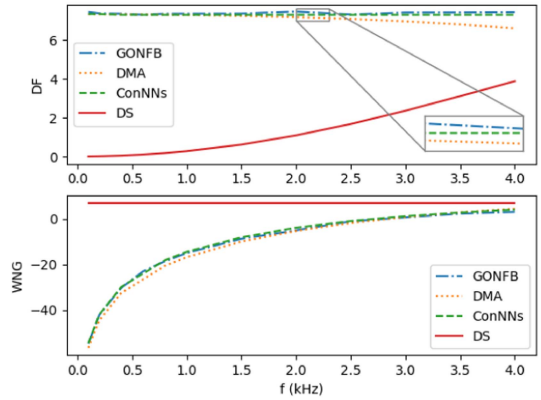



Fig. 8. Performance of GONFB, DMA, ConNNs and DS in a circular array when $M = 5$ and radius $r = 0.015$ m.

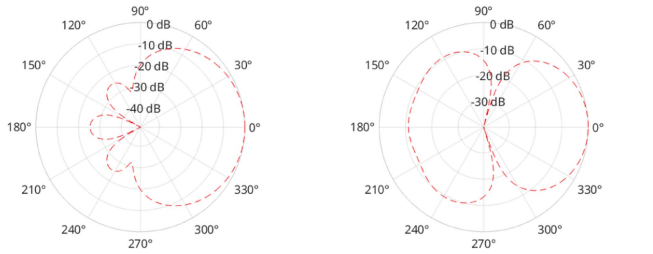
achieves a more effective linear array geometry along with its corresponding filter design.

### C. Circular Array

*1) Initialisation and Hyperparameters:* The circular array to be optimised had 5 microphones. The radius of the array was 1.5 cm. The initial filter was obtained in the same way as the linear array. The geometry of the array was fixed to a uniform circular array. Thus, we only train FNet in this scenario, which had 15 hidden layers. Each layer had 40 neurons. The loss was evaluated every 100,000 epochs. The parameters of the augmented Lagrangian were updated every 40,000 epochs. The maximum epoch number was 1,000,000. The target DF value was 7.4 dB. With a tolerance of 0.1 dB, we set $\alpha_i = 7.3$ and $\beta_i = 7.5$. The frequency range was from 100 Hz to 4000 Hz, since the DMA-based approach we compared with [20] could only be frequency invariant below 4000 Hz. The direction of the incoming signal was $0^\circ$. The baseline DMA method was from [20].

*2) Performance Analysis:* The performance comparison of our circular array with [20], DS, and ConNNs is shown in Fig. 8. With the same geometry, both GONFB and ConNNs exhibit superior performance in DF values, while our WNG values are close to or slightly better than the WNGs in [20]. The DF values of DS are significantly lower than other approaches. Fig. 8 shows that both GONFB and ConNNs can optimise the spatial filter effectively when the geometry is fixed.

We present a comparison of UCMA beampatterns in Fig. 9. The DMA-based beampattern utilises the Jacobi-Anger series expansion technique to approximate the second-order supercardioid pattern, whereas the GONFB-based beampattern is automatically designed through neural networks. It is evident that the GONFB-based beampattern features a single null direction and is significantly simpler compared to the DMA-based beampattern. However, despite its simplicity, the GONFB-based beampattern exhibits superior directivity and robustness. This observation strengthens our findings that manually designed null directions and differential orders by human experts are




(i) DMA (ii) GONFB

Fig. 9. Comparison of UCMA beampatterns when $M = 5, r = 0.015$ m, $f = 4$ kHz. (i) The beampattern designed by DMA. (ii) The beampattern designed by GONFB.

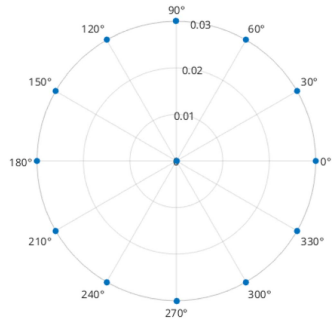



Fig. 10. optimised geometry from GONFB for a uniform concentric circular array. 12 microphones in one ring and 1 additional microphone in the center. Microphones are represented by blue circles.

suboptimal compared to the beampattern parameters optimised by neural networks.

### D. Concentric Circular Array

*1) Initialisation and Hyperparameters:* The concentric circular array to be optimised had 13 microphones in total. The initial geometry of the array consisted of two rings. Each ring had 6 uniformly located microphones. An additional microphone was located at the center of the array. The minimum radius of the inner ring was 1 cm and the maximum radius of the outer ring was 3 cm. The initial filter was obtained by applying the null-constrained approach, with null directions in $102^\circ$ and $174^\circ$. Before training, the angles of the starting microphones on the inner and outer rings were $30^\circ$ and $0^\circ$, respectively. The initial radii of the inner and outer rings were 2cm and 3cm, respectively. In the GNet, the radius of each ring and the angular position of the first microphone on each ring were optimised. FNet had 10 hidden layers. Each layer had 90 neurons. The loss was evaluated every 80,000 epochs. Parameters of the augmented Lagrangian were updated every 10,000 epochs. The maximum epoch number was 600,000. The target DF value was 8.67 dB. $\alpha_i$ was set to the corresponding DF value of [26]. $\beta_i = 8.67 + |8.67 - \alpha_i|$. The frequency range was from 100 Hz to 8000 Hz. The direction of the incoming signal was $30^\circ$. The baseline DMA-CCMA-II method was from [26].

*2) Performance Analysis:* The optimised geometry of the concentric circular array from GONFB is shown in Fig. 10. Contrary to the research finding in [26], our approach prefers to use only one ring instead of two rings. In other words, placing 12 microphones all in one ring can perform better than dividing the microphones into two rings. The superior performance of two rings in [26] may be affected by the fact that the authors use six additional microphones in two rings. ConNNs fail to optimise the geometry of the concentric circular array and there is almost no change from the initial geometric configuration.

We present a performance analysis of our UCCMA technique in comparison with the DMA-CCMA-II method utilising series expansions as detailed in [26], along with ConNNs and DS. Throughout this evaluation, all beamforming strategies except DS are required to maintain frequency invariance.

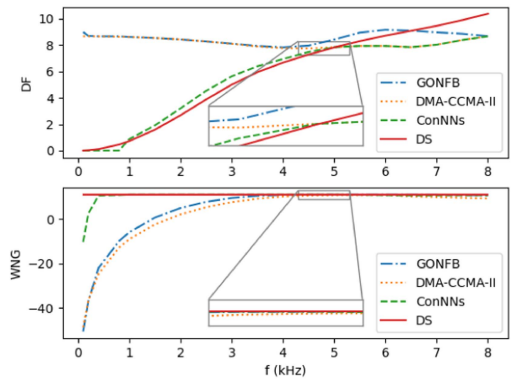



Fig. 11. Performance of GONFB, DMA-CCMA-II, ConNNs and DS in a concentric circular array when $M = 13$ and the maximum radius $r_{\max} = 0.03$ m.

The results of the performance evaluation are shown in Fig. 11. ConNNs and DS exhibit similar performance; while both maintain high WNG values across frequencies, they fail to construct a frequency-invariant fixed beamformer, as reflected by their consistently low DF values, particularly below 5000 Hz. In contrast, GONFB and DMA-CCMA-II successfully achieve frequency-invariant behavior. Notably, GONFB demonstrates superior DF values at frequencies above 4000 Hz and also achieves higher WNG values. This further underscores the robustness and effectiveness of the GONFB technique across diverse geometric scenarios.

In our comparative analysis, we assess the performance of UCMA and UCCMA when optimised using GONFB. The angle of the starting microphone is $30^\circ$, and we evaluate the loss every 100,000 epochs, updating the parameters of the augmented Lagrangian every 40,000 epochs, with a maximum of 2,000,000 epochs. Each layer has 180 neurons. Other details of network training are identical to those of UCCMA. For a fair comparison, the radius of the CMA is equal to the maximum radius of the outer ring in the UCCMA, measuring 3cm. Both UCMA and UCCMA employ 13 microphones. However, the key distinction lies in the placement: in UCCMA, one microphone is centrally located, while UCMA positions all microphones along its ring. Our findings, illustrated in Fig. 12, demonstrate that UCMA, when optimised through GONFB, not only resolves deep

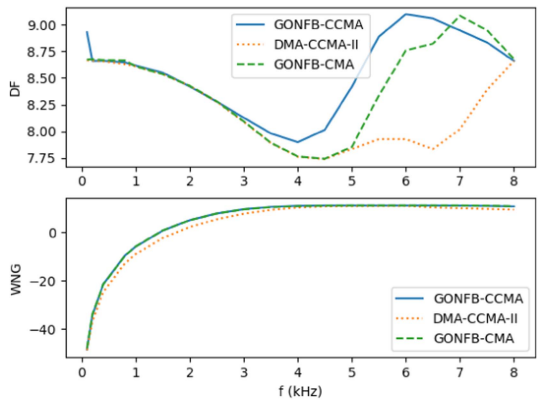



Fig. 12. Performance of GONFB-CCMA, DMA-CCMA-II, and GONFB-CMA when $M = 13$. The radius of of GONFB-CMA is 3 cm, which is the same as the maximum radius of the outer ring in GONFB-CCMA.

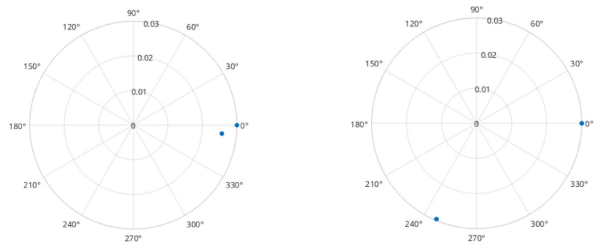



Fig. 13. Optimised Geometry of two microphones when maximizing only DF. The look direction is 30°. Microphones are represented by blue circles. (i) The array geometry designed by GONFB with minimum microphone distance. (ii) The array geometry designed by ConNNs is not optimal.

null issues but also delivers comparable performance to UC-CMA in both GONFB and DMA configurations. This suggests that UCMA's geometric configuration could offer competitive beamforming performance.

### *E. Circular Discal Array*

*1) Initialisation and Hyperparameters:* To explore the behavior of GONFB in optimising a circular discal microphone array, we first considered two straightforward scenarios: maximising only WNG or maximising only DF with two microphones. The minimum spacing between the microphones was set at 0.5 cm. The radius of the disc was 3 cm. The initial filter was randomly sampled from a uniform distribution. The initial locations of the two microphones were at the intersections between the rim of the disc and the $x$-axis. To avoid geometric degeneration when maximising DF, the microphone location at $0^\circ$ is fixed. FNet had 10 hidden layers. Each layer had 90 neurons. The loss was evaluated every 80,000 epochs. The maximum epoch number was 800,000. The frequency considered was 1000 Hz. As only one frequency value was considered, we did not use the augmented Lagrangian in the loss function. Instead, the loss function maximized (8) or (9). Additionally, a penalty term was appended in the loss function to ensure the minimum distance between the two microphones was not violated. The direction of the incoming signal was $30^\circ$.

*2) Performance Analysis:* Two microphones are invariably configured into a linear array on a disc. The maximum WNG of a distortionless linear array is equal to the number of microphones [21]. Both GONFB and ConNNs effectively determine the global optimum, yielding a WNG value of 2. Since (8) does not rely on geometric information for maximization, the geometric configurations are degenerate. In simpler terms, numerous viable geometric arrangements for a linear array on a disc can achieve the maximum WNG, provided the filter requirement is satisfied. The filter for optimal WNG can be analytically deduced as:

$$\mathbf{h}_{\text{MWNG}}(\omega) = \mathbf{d}_{\text{CDA}}(\omega)/M. \tag{39}$$

The observation that ConNNs can optimise a competitive CDA spatial filter is consistent with our observations in the CMA scenario. When geometric optimisation is dispensable, ConNNs demonstrate an aptitude to optimise spatial filters.

The maximum DF of a distortionless linear array can be analytically derived, as shown in [21]. The expression for the maximum DF, denoted as $\text{DF}[\mathbf{h}(\omega)]_{\max}$, is given by:

$$\text{DF}[\mathbf{h}(\omega)]_{\max} = 2\frac{1-\text{sinc}(\omega\tau_0)\cos(\omega\tau_0\cos(\phi-\phi_{\text{CDA}}))}{1-\text{sinc}^2(\omega\tau_0)}, \tag{40}$$

where $\tau_0$ represents the distance $l_0$ between two microphones divided by $c$, and $\phi_{\text{CDA}}$ is the azimuth angle of the linear microphone array formed by these two microphones. Equation (40) highlights that for optimal performance, the discrepancy between angles $\phi$ and $\phi_{\text{CDA}}$ should ideally be zero, and the distance $l_0$ should be minimized. The arrangement illustrated in Fig. 13 demonstrates that the GONFB-based array aligns precisely with the incoming signal direction. The inter-microphone spacing is optimised to the minimum distance. The DF value of the optimised array aligns numerically with the theoretical maximum DF value derived from Equation (40), resulting in a value of 6.02dB. Evidently, GONFB has effectively located the global optimum in this scenario. In contrast, ConNNs exhibit a tendency to converge to local optima and are unable to effectively minimize microphone spacing. The azimuth angle $\phi_{\text{CDA}}$ produced by ConNNs is approximately $33^\circ$, slightly deviating from the desired direction of $30^\circ$ . Consequently, the optimised directivity factor for ConNNs is recorded at 5.76dB.

Furthermore, we compared a CDA with an LA following the same parameter settings in V-B. In Fig. 14, we show that a CDA designed by GONFB outperforms its LA counterpart in terms of WNGs. By placing 16 microphones in a novel geometry as illustrated in Fig. 15, the CDA achieves comparable DF values to those of the LA while maintaining frequency invariance. The CDA also outperforms the LA by significantly enhancing the WNGs across all frequency bins. After 1 kHz, the WNGs of the CDA are about 10 dB higher than those of the LA. This comparison demonstrates the significance of geometric optimisation and the potential of the CDA geometry.

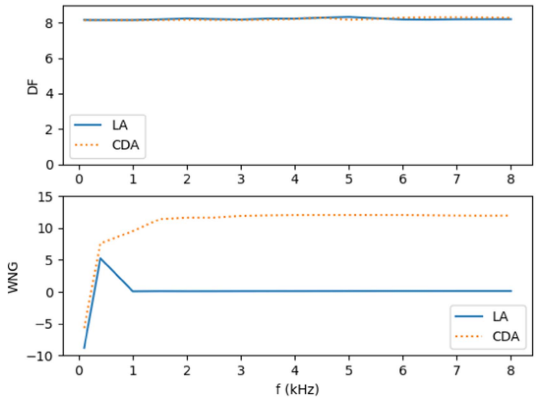



Fig. 14. Performance of LA vs CDA when both arrays are designed by GONFB. $M = 16$, $L = 0.15$ m and $r = 0.075$ m.

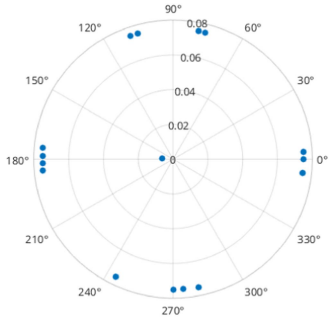
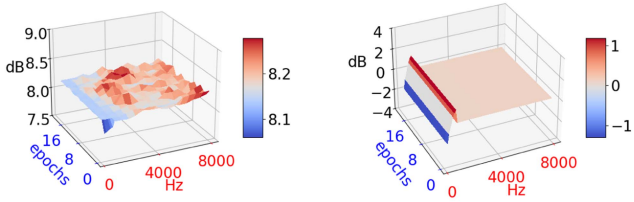


Fig. 15. Optimised circular discal array geometry of GONFB when $M = 16$ and $r = 0.075$ m. The look direction is end-fire.




(i) DF (ii) WNG

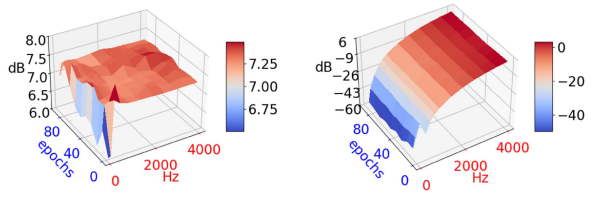
Fig. 16. The convergence plot of DF and WNG values in a linear array during the training of GONFB.

## VI. Further Analysis

This section presents a convergence analysis of the GONFB training process for three array configurations: linear, circular, and concentric circular arrays. Additionally, the computational complexity of GONFB is analysed for each of these geometric setups. The circular discal array, being primarily an exploratory case with a simplified GONFB implementation, is excluded from the convergence and complexity analysis in this section.

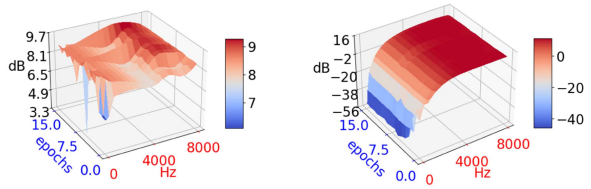
### A. Convergence Analysis

In Fig. 16, 17 and 18, we present the convergence plots of DF and WNG values during the GONFB training for linear, circular, and concentric circular arrays. The horizontal axes represent the




(i) DF (ii) WNG

Fig. 17. The convergence plot of DF and WNG values in a circular array during the training of GONFB.




(i) DF (ii) WNG

Fig. 18. The convergence plot of DF and WNG values in a concentric circular array during the training of GONFB.

frequency bins and the number of epochs. Note that the starting point at 0 on the epochs axis does not correspond to epoch 0, but rather to the first recorded epoch during the training process. Each epoch unit represents 10,000 epochs. The vertical axis maps the DF and WNG values in decibels, with darker blue indicating lower values and darker red signifying higher values. The training and geometric setups follow those described in Section V-B,V-C, and V-D.

Across all geometric configurations, WNG values exhibit small variation at the beginning of training, with the remainder of the process primarily fine-tuning these values. In contrast, DF values undergo significantly larger changes during convergence, which is expected given the optimisation process of our loss function using augmented Lagrangian methods, as described in Section IV-C. The DF values are dynamically constrained by $c_k^i$ as shown in (38), allowing for considerable fluctuation as training progresses.

In Fig. 16, we observe that the DF values are initially too low at low-frequency bins and too high at high-frequency bins. Over time, these extreme values are smoothed, leading to more frequency-invariant DF values. By the final epochs, the DF values have flattened significantly compared to the initial stages. In Fig. 17, DF convergence is especially noticeable in low-frequency bins below 400 Hz. A clear pattern emerges where the low DF values are progressively elevated to match the higher-frequency bins as training goes. In Fig. 18, we not only observe the enhancement of DF values at low frequencies but also the presence of multiple valleys in DF values over the epochs. This is consistent with the optimisation process in the concentric circular geometry. When the geometry undergoes major changes, such as two rings merging into one, the DF values initially drop due to the structural change but quickly adapt and improve in the new configuration.

In summary, both DF and WNG values converge during GONFB training. The DF values exhibit substantial changes

TABLE I
COMPLEXITY COMPARISON OF GONFB ACROSS DIFFERENT GEOMETRIC SETTINGS: LINEAR, CIRCULAR, AND CONCENTRIC CIRCULAR ARRAYS

| | GNet | | FNet | |
|---|---|---|---|---|
| Geometry | #Params | #FLOPs | #Params | #FLOPs |
| LA | 3.8 | 3.8 | 66.4 | 2638.1 |
| CMA | N/A | N/A | 5 | 117.1 |
| CCMA | 0.9 | 0.9 | 16.9 | 633.4 |

The number of parameters (#Params) and floating-point operations (#FLOPs) in both GNet and FNet are compared. All values are rounded to one decimal place and are presented in units of $\times 10^4$.

throughout the process, driven by the requirement for frequency invariance and the augmented Lagrangian optimisation. In contrast, WNG values display a smoother, more gradual convergence pattern.

### B. Complexity Analysis

When analyzing the complexity of neural networks, the number of parameters (#Params) and floating-point operations (#FLOPs) are commonly used metrics [65], [66]. The #Params measure reflects the memory footprint, while the #FLOPs metric indicates computational cost. Higher values of #Params and #FLOPs suggest greater complexity.

The details of #Params and #FLOPs in GONFB across different geometric configurations are presented in Table I. For GNet comparison, CMA is excluded as it does not require geometry optimisation. The GNet in LA has higher #Params and #FLOPs than in CCMA. This is because, in LA, geometry optimisation involves adjusting the spacing between all microphones, while in CCMA, GNet focuses only on optimising the radii and the starting angular positions of the two rings. As a result, GNet in LA handles more optimisation tasks and thus requires greater complexity.

For FNet comparison, CMA has the fewest #Params and #FLOPs, which can be attributed to the smaller number of microphones ($M = 5$) and the restricted frequency range (up to 4000 Hz) considered in this configuration. In contrast, both LA and CCMA need to account for a frequency range up to 8000 Hz. In LA, 16 microphones are used, all with optimisable positions. In CCMA, although there are 13 microphones, one is fixed, and the remaining 12 are divided into two groups with fixed geometry within each group. This makes the filtering design task in CCMA less challenging than in LA.

In summary, the complexity of GONFB aligns with the difficulty of the optimization tasks associated with each geometric configuration. The complexity of GNet is driven by the number of parameters and locations to optimise, while the complexity of FNet depends on the frequency range, the number of microphones, and the complexity of geometric optimisation.

## VII. CONCLUSION

This paper presented a novel and comprehensive framework for the neural optimisation of fixed beamformers with varying geometries. The proposed neural network model, Geometrically optimised Neural Fixed Beamformer (GONFB), successfully optimised both array geometries and spatial filters of fixed beamformers. By leveraging the ResNet structure and incorporating an augmented Lagrangian-based loss function, GONFB surpassed its DMA-based counterparts in linear, circular, and concentric circular arrays. GONFB also outperformed Constrained Naked Neurons (ConNNs) in optimising array geometries.

In our study, we observed consistent superior performance of GONFB in terms of DF and WNG across different frequency bins. GONFB exhibited these advantages while maintaining a desirable frequency-invariant property. Our experiments revealed that GONFB could effectively design CCMAs using a single ring. Moreover, our results demonstrated that CMAs designed by GONFB performed comparably to CCMAs. Additionally, we explored the capabilities of GONFB in circular discal geometric settings and achieved globally optimal solutions using two microphones. When equipped with an equal number of microphones, a circular discal array can outperform a linear array due to its superior robustness. These results showcase the versatility and robustness of GONFB in various array configurations, emphasising its potential in diverse applications. Moreover, they demonstrate GONFB's potential for practical applications and further research in array signal processing.

## ACKNOWLEDGMENT

This work was done while Weilong Huang was a Senior Algorithm Engineer with Alibaba Group, Hangzhou, China. The research was partly conducted while Longfei Felix Yan was a dual PhD student in Victoria University of Wellington and the Australian National University.

## REFERENCES

[1] G. W. Elko, "Microphone array systems for hands-free telecommunication," *Speech Commun.*, vol. 20, no. 3/4, pp. 229–240, 1996.
[2] W. Herbordt, T. Horiuchi, M. Fujimoto, T. Jitsuhiro, and S. Nakamura, "Hands-free speech recognition and communication on PDAs using microphone array technology," in *Proc. IEEE Workshop Autom. Speech Recognit. Understanding*, IEEE, 2005, pp. 302–307.
[3] I. I. Papp, Z. M. Saric, and N. D. Teslic, "Hands-free voice communication with TV," *IEEE Trans. Consum. Electron.*, vol. 57, no. 2, pp. 606–614, May 2011.
[4] K. Kumatani, J. McDonough, and B. Raj, "Microphone array processing for distant speech recognition: From close-talking microphones to far-field sensors," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 127–140, Nov. 2012.
[5] W. Jin, M. J. Taghizadeh, K. Chen, and W. Xiao, "Multi-channel noise reduction for hands-free voice communication on mobile phones," in *Proc. 2017 IEEE Int. Conf. Acoust., Speech Signal Process.*, 2017, pp. 506–510.
[6] J. Chua and W. B. Kleijn, "A low latency approach for blind source separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 8, pp. 1280–1294, Aug. 2019.
[7] H. Sun, P. Samarasinghe, and T. Abhayapala, "Blind source counting and separation with relative harmonic coefficients," in *Proc. 2023 IEEE Int. Conf. Acoust., Speech Signal Process.*, 2023, pp. 1–5.
[8] Y. Hu, W. Wang, Z. Gu, T. Mao, X. Zhu, and J. Jin, "Closed-form multiple source direction-of-arrival estimator under reverberant environments," *J. Acoustical Soc. Amer.*, vol. 154, no. 4, pp. 2349–2364, 2023.
[9] W. Manamperi, T. D. Abhayapala, J. Zhang, and P. N. Samarasinghe, "Drone audition: Sound source localization using on-board microphones," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 508–519, 2022.
[10] S. K. Yadav and N. V. George, "Joint dereverberation and beamforming with blind estimation of the shape parameter of the desired source prior," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 32, pp. 779–793, 2024.

[11] W. N. Manamperi, T. D. Abhayapala, P. N. Samarasinghe, and J. A. Zhang, "Drone audition: Audio signal enhancement from drone embedded microphones using multichannel wiener filtering and Gaussian-mixture based post-filtering," *Appl. Acoust.*, vol. 216, 2024, Art. no. 109818.

[12] H. N. Chau, T. D. Bui, H. B. Nguyen, T. T. Duong, and Q. C. Nguyen, "A novel approach to multi-channel speech enhancement based on graph neural networks," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 32, pp. 1133–1144, 2024.

[13] J. Benesty, J. Chen, and C. Pan, *Fundamentals of Differential Beamforming*. Singapore: Springer, 2016.

[14] G. Huang, J. Benesty, and J. Chen, "Fundamental approaches to robust differential beamforming with high directivity factors," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 3074–3088, 2022.

[15] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[16] G. W. Elko, "Superdirectional microphone arrays," in *Acoustic Signal Processing for Telecommunication*. Boston, MA, USA: Springer, 2000, pp. 181–237.

[17] J. Benesty and J. Chen, *Study and Design of Differential Microphone Arrays*, vol. 6. Berlin, Germany: Springer, 2012.

[18] J. Chen, J. Benesty, and C. Pan, "On the design and implementation of linear differential microphone arrays," *J. Acoustical Soc. America*, vol. 136, no. 6, pp. 3097–3113, 2014.

[19] Y. Buchris, I. Cohen, and J. Benesty, "First-order differential microphone arrays from a time-domain broadband perspective," in *Proc. 2016 IEEE Int. Workshop Acoust. Signal Enhancement*, 2016, pp. 1–5.

[20] G. Huang, J. Benesty, and J. Chen, "On the design of frequency-invariant beampatterns with uniform circular microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 5, pp. 1140–1153, May 2017.

[21] J. Benesty, I. Cohen, and J. Chen, *Fundamentals of Signal Enhancement and Array Signal Processing*. Hoboken, NJ, USA: Wiley, 2017.

[22] F. Borra, A. Bernardini, F. Antonacci, and A. Sarti, "Uniform linear arrays of first-order steerable differential microphones," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 12, pp. 1906–1918, Dec. 2019.

[23] H. Zhang, J. Chen, and J. Benesty, "Study of nonuniform linear differential microphone arrays with the minimum-norm filter," *Appl. Acoust.*, vol. 98, pp. 62–69, 2015.

[24] J. Jin, J. Benesty, G. Huang, and J. Chen, "On differential beamforming with nonuniform linear microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 1840–1852, 2022.

[25] J. Benesty, J. Chen, and I. Cohen, *Design of Circular Differential Microphone Arrays*, vol. 12. Berlin, Germany: Springer, 2015.

[26] X. Wang, G. Huang, I. Cohen, J. Benesty, and J. Chen, "Robust steerable differential beamformers with null constraints for concentric circular microphone arrays," in *Proc. ICASSP 2021—2021 IEEE Int. Conf. Acoust., Speech Signal Process.*, 2021, pp. 4465–4469.

[27] G. Huang, J. Benesty, I. Cohen, and J. Chen, "A simple theory and new method of differential beamforming with uniform linear microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 1079–1093, 2020.

[28] L. Yan, W. Huang, W. B. Kleijn, and T. D. Abhayapala, "Phase error analysis for first-order linear differential microphone arrays," in *Proc. 2022 Int. Workshop Acoust. Signal Enhancement*, 2022, pp. 1–5.

[29] C. Pan, J. Chen, and J. Benesty, "Theoretical analysis of differential microphone array beamforming and an improved solution," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 11, pp. 2093–2105, Nov. 2015.

[30] W. Huang and J. Feng, "Differential beamforming for uniform circular array with directional microphones," in *INTERSPEECH*, 2020, pp. 71–75.

[31] W. Huang and J. Feng, "Minimum-norm differential beamforming for linear array with directional microphones," in *Proc. Interspeech*, 2021, pp. 701–705.

[32] W. Huang and J. Feng, "Robust steerable differential beamformer for concentric circular array with directional microphones," in *Proc. 2022 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2022, pp. 319–323.

[33] X. Luo, J. Jin, G. Huang, J. Chen, and J. Benesty, "Design of steerable linear differential microphone arrays with omnidirectional and bidirectional sensors," *IEEE Signal Process. Lett.*, vol. 30, pp. 463–467, 2023.

[34] X. Luo et al., "Constrained maximum directivity beamformers based on uniform linear acoustic vector sensor arrays," in *Proc. 2021 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2021, pp. 1221–1225.

[35] X. Luo et al., "Design of maximum directivity beamformers with linear acoustic vector sensor arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 1421–1435, 2023.

[36] S. J. Patel, S. L. Grant, M. Zawodniok, and J. Benesty, "On the design of optimal linear microphone array geometries," in *Proc. 16th Int. Workshop Acoust. Signal Enhancement*, 2018, pp. 501–505.

[37] J. Jin, G. Huang, J. Chen, and J. Benesty, "Design of optimal linear differential microphone arrays based array geometry optimization," in *Proc. 2019 IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 5741–5745.

[38] Y. Konforti, I. Cohen, and B. Berdugo, "Array geometry optimization for region-of-interest broadband beamforming," in *Proc. 2022 Int. Workshop Acoust. Signal Enhancement*, 2022, pp. 1–5.

[39] R. Moisseev, G. Itzhak, and I. Cohen, "Array geometry optimization for region-of-interest near-field beamforming," in *Proc. ICASSP 2024-2024 IEEE Int. Conf. Acoust., Speech Signal Process.*, 2024, pp. 576–580.

[40] L. Yan, W. Huang, W. B. Kleijn, and T. D. Abhayapala, "Neural optimization of geometry and fixed beamformer for linear microphone arrays," in *Proc. ICASSP 2023—2023 IEEE Int. Conf. Acoust., Speech Signal Process.*, 2023, pp. 1–5.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[42] Z. Ding, S. Chen, Q. Li, and S. Wright, "On the global convergence of gradient descent for multi-layer resnets in the mean-field regime," 2021, *arXiv:2110.02926*.

[43] Z. Ding, S. Chen, Q. Li, and S. J. Wright, "Overparameterization of deep resnet: Zero loss and mean-field analysis," *J. Mach. Learn. Res.*, vol. 23, no. 48, pp. 1–65, 2022.

[44] R. DeVore, B. Hanin, and G. Petrova, "Neural network approximation," *Acta Numerica*, vol. 30, pp. 327–444, 2021.

[45] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Hoboken, NJ, USA: John Wiley & Sons, 2002.

[46] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, vol. 1. Berlin, Germany: Springer, 2008.

[47] E. Gilbert and S. Morgan, "Optimum design of directive antenna arrays subject to random variations," *Bell Syst. Tech. J.*, vol. 34, no. 3, pp. 637–663, 1955.

[48] J. Wang, F. Yang, and J. Yang, "A perspective on fully steerable differential beamformers for circular arrays," *IEEE Signal Process. Lett.*, vol. 30, pp. 648–652, 2023.

[49] G. Huang, J. Benesty, and J. Chen, "Design of robust concentric circular differential microphone arrays," *J. Acoustical Soc. Amer.*, vol. 141, no. 5, pp. 3236–3249, 2017.

[50] D. Yarotsky, "Error bounds for approximations with deep ReLU networks," *Neural Netw.*, vol. 94, pp. 103–114, 2017.

[51] M. I. Belghazi et al., "Mutual information neural estimation," in *Proc. 35th Int. Conf. Mach. Learn.*, Jul. 2018, vol. 80, pp. 531–540.

[52] L. Yan, W. B. Kleijn, and T. Abhayapala, "A linear-time independence criterion based on a finite basis approximation," in *Proc. 23rd Int. Conf. Artif. Intell. Statist.*, Aug. 2020, vol. 108, pp. 202–212.

[53] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[54] M. W. Matthès, Y. Bromberg, J. de Rosny, and S. M. Popoff, "Learning and avoiding disorder in multimode fibers," *Phys. Rev. X*, vol. 11, no. 2, 2021, Art. no. 021060.

[55] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUS)," 2016, *arXiv:1606.08415*.

[56] M.-H. Guo et al., "Attention mechanisms in computer vision: A survey," *Comput. Vis. Media*, vol. 8, no. 3, pp. 331–368, 2022.

[57] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI Open*, vol. 3, pp. 111–132, 2022.

[58] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 28492–28518.

[59] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1606.08415*.

[60] M. J. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*. New York, NY, USA: Academic Press, 1969, pp. 283–298.

[61] S. Wright and J. Nocedal, *Numerical Optimization*, vol. 35. New York, NY, USA: Springer, 1999.

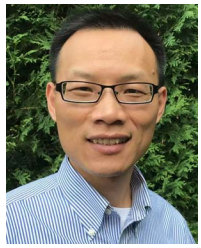
[62] D. P. Bertsekas et al. *Constrained Optimization and Lagrange Multiplier Methods*. Belmont, MA, USA: Athena Scientific, 1996.

[63] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, San Diego, CA, USA, May 2015.
[64] R. A. Kennedy, T. D. Abhayapala, and D. B. Ward, "Broadband nearfield beamforming using a radial beampattern transformation," *IEEE Trans. Signal Process.*, vol. 46, no. 8, pp. 2147–2156, Aug. 1998.
[65] L. Yang et al., "CondenseNet V2: Sparse feature reactivation for deep networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 3569–3578.
[66] J. Huang, B. Xue, Y. Sun, M. Zhang, and G. G. Yen, "Split-level evolutionary neural architecture search with elite weight inheritance," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 10, pp. 13523–13537, Oct. 2023.

**Longfei Felix Yan** (Member, IEEE) received the B.Sc. (with Hons.) degree from the Victoria University of Wellington (VUW), Wellington, New Zealand, in 2017, and the dual Ph.D. degree from VUW and Australian National University, Canberra, ACT, Australia, in 2024. He is currently a Lecturer with VUW. His research interests include statistical signal processing, machine learning, and combinatorial optimisation.

**Weilong Huang** received the B.Sc. degree from Shandong University, Jinan, China, in 2011, and the M.Sc. degree from Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, in 2014. He is currently working toward the Doctoral degree with International Audio Laboratories Erlangen (a Joint Institution of Friedrich-Alexander-Universität Erlangen-Nürnberg and Fraunhofer IIS, Germany). From 2016 to 2019, he was a Research Engineer with Netease Inc., China. From 2019 to 2024, he was a Senior Algorithm Engineer (Algorithm Expert) with Alibaba Group, China. His research interests include speech signal processing and machine learning algorithms with a focus on multichannel speech enhancement. He was the recipient of the 13 granted patents with Alibaba Group.

**Thushara D. Abhayapala** (Fellow, IEEE) received the B.E. degree in engineering and the Ph.D. degree in telecommunication engineering from the Australian National University (ANU), Canberra, ACT, Australia, in 1994 and 1999, respectively. He is currently a Professor of audio and acoustic signal processing with ANU. From 2015 to 2019, he was the Deputy Dean with the ANU College of Engineering and Computer Science, Head with the ANU Research School of Engineering from 2010 to 2014, and a Leader with Wireless Signal Processing Program, National ICT Australia, Australia, from 2005 to 2007. He has supervised 44 Ph.D. degree students and has coauthored more than 300 peer-reviewed papers. His research interests include the areas of spatial audio and acoustic signal processing, and multichannel signal processing. Among many contributions, he is one of the First Researchers to use spherical harmonic based Eigen-decomposition in microphone arrays and to propose the concept of spherical microphone arrays, and was one of the first to show the fundamental limits of spatial sound field reproduction using arrays of loudspeakers and spherical harmonics which is now termed as higher order Ambisonics. He also made fundamental contributions to the problem of the multizone sound field reproduction. He worked in industry for two years, before his doctoral study and has active collaboration with a number of companies. He was the Co-Chair of IEEE WASPAA 2021. He was an Associate Editor for IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING. From 2011 to 2016, he was a Member of the Audio and Acoustic Signal Processing Technical Committee of the IEEE Signal Processing Society. He is a Fellow of Engineers Australia.



**Jinwei Feng** received the Bachelor of Science and Master of Science degrees from the Department of Electronic Science, Nanjing University, Nanjing, China, in 1992 and 1995, respectively, the Master of Engineering degree in signal processing from Nanyang Technical University, Singapore in 1997, and the Ph.D. degree in acoustics from Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, in 2000. He has authored about 40 patents and 30 research papers. His research include acoustic design and audio signal processing for audio and video conferencing applications. Dr. Feng was the Principal investigator in putting out the world's first commercially successful voice-tracking smart camera. The innovation has thereafter been imitated by all major players in the industry which have created a multibillion-dollar market.

**W. Bastiaan Kleijn** (Fellow, IEEE) received the Ph.D. degree in soil science, the M.Sc. degree in physics from the University of California, Riverside, CA, USA, the M.S.E.E. degree from Stanford University, Stanford, CA, USA, and the Ph.D. degree in electrical engineering from TU Delft, Delft, Netherlands. He was a Member of Technical Stafff in the Research Division, AT&T Bell Laboratories. Since 2010, he has been a Professor with the Victoria University of Wellington, Wellington, New Zealand and has also been a Research Scientist with Google, since 2011. From 2011 to 2021, he was a Professor with TU Delft and he also was a Professor at KTH Stockholm from 1996 until 2014. In 2021. He is a Fellow of the Royal Society of New Zealand and a Fellow of Engineering New Zealand.