Deep Learning-Based Prediction of Far-Field Sound Directivity in Parametric Array Loudspeakers

Shangming Mei

PhD

University of York

Physics, Engineering and Technology

June 2025

Abstract

Parametric array loudspeakers (PALs) enable highly directional sound projection through ultrasonic modulation and are widely used in applications such as immersive audio, assistive communication, and spatial sound control. Despite their potential, accurate far-field directivity modelling remains challenging due to nonlinear acoustic propagation and the complexity of sparse array configurations. This dissertation proposes an integrated framework combining analytical, numerical, and deep learning methods to address these challenges. A dual convolution model is developed to improve grating lobe prediction and nonlinear beam characterisation. To optimise array performance under varying steering angles, a particle swarm optimisation (PSO) method is introduced for transducer layout design. Full-wave acoustic simulations using the finite-element method (FEM) are employed to model propagation and validate the optimised arrays, demonstrating up to a 5 dB reduction in peak sidelobe level across the tested steering range. To overcome the computational burden of full-wave simulations, deep learning models, based on generative adversarial network (GAN) architectures are trained to infer far-field patterns from sparse near-field inputs. By integrating physics-informed modelling, numerical optimisation, and data-driven prediction, the proposed approach improves both design efficiency and accuracy. In particular, the GAN-based predictor achieves MAE = 0.0105 and RMSE = 0.0171 on peak-normalised directivity. Experimental measurements were conducted to validate the proposed models and inform their development, ensuring consistency between predicted and observed acoustic directivity patterns. These outcomes address critical limitations in existing PAL modelling techniques and support the feasibility of real-time, adaptive control of acoustic fields for advanced sound applications.

Key words: Parametric Array Loudspeaker; Far-field Directivity; Dual Convolution Model; Generative Adversarial Network; Finite Element Method; Acoustic Beamforming; Nonlinear Acoustics; Spatial Sound Field Prediction

Table of Contents

Abbreviations

2D	Two-dimensional
3D	Three-dimensional
AM	Amplitude modulation
ANC	Active noise control
CMOS	Complementary metal oxide semiconductor
CMUTs	Capacitive micromachined ultrasonic transducers
dB	Decibels
DSB	Double sideband
DSP	Digital signal processing
DOF	Degrees of freedom
FDM	Finite-difference method
FEM	Finite element method
FFT	Fast Fourier Transform
FM	Frequency modulation
GA	Genetic algorithm
GAN	Generative adversarial network
GaN	Gallium nitride
HIFU	High-intensity focused ultrasound
IQR	Interquartile range
KZK	Khokhlov-Zabolotskaya-Kuznetsov
MAE	Mean absolute error
MAM	Modified amplitude modulation
PAL	Parametric array loudspeaker
PDM	Pulse density modulation
PMUTs	Piezoelectric micromachined ultrasonic transducers

PSL	Peak side lobe level
PSO	Particle swarm optimisation
PWM	Pulse width modulation
PML	Perfectly Matched Layer
QPSO	Quantum particle swarm optimisation
RH	Relative humidity
RAM	Recursive amplitude modulation
RASS	Radio acoustic sounding systems
RMSE	Root Mean Squared Error
SiC	Silicon carbide
SNR	Signal-to-noise ratio
SOS	Symbiotic organisms search
SPL	Sound pressure level
SRAM	Square-Root amplitude modulation
SSB	Single-sideband modulation
SSB-WC	Single sideband with carrier
THD	Total harmonic distortion
THD+N	Total harmonic distortion plus noise
WGANs	Wasserstein GANs

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, Dr. Yihua Hu and Dr. Mohammad Nasr Esfahani, for their unwavering support, insightful guidance, and continuous encouragement throughout my PhD journey. Their mentorship has shaped me not only as a researcher but also as a more thoughtful and resilient individual. I am deeply honoured to have been their student and hope my work has lived up to their expectations.

I am profoundly thankful to my parents for their unconditional love, patience, and belief in me. Their support has been my greatest source of strength and inspiration.

My sincere thanks also go to Dr. John Dowson and Dr. Xing Zhao for their valuable suggestions and constructive feedback on my research.

I am especially grateful to my colleagues and friends Dr. Wangjie Lang, Dr. Xiaotian Zhang, Dr. Hongzuo Liu, Dr. Nie Yu, and Mr. Mengyu Cheng for their stimulating discussions, generous help, and many shared moments of joy.

I gratefully acknowledge the support of the School of Physics, Engineering and Technology at the University of York. Special thanks to Ms. Camilla Danese and Ms. Jenni Pickin for their patience, kindness, and administrative assistance throughout my time at York.

Finally, my sincere thanks to everyone who helped me reach this milestone.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

Content from **Chapter 2** has been published in *Springer Proceedings* and *IEEE Access* under the following titles: “The Parametric Array Speaker: A Review” (Springer, 2022)” and “The Class D Audio Power Amplifier: A Review” (Electronics, 2022)

Content from **Chapter 3** is based on the same IEEE Access publication with the following title: “A Dual Convolution Model for Grating Lobe Prediction in Directivity of Parametric Array Loudspeakers” (*IEEE Access*, 2025)

Content from **Chapter 4** has been submitted to *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, under the title: “Adaptive Optimisation of Non-Uniform Phased Array Speakers Using Particle Swarm Optimisation for Enhanced Directivity Control.” The manuscript is currently under review.

Content from **Chapter 5** has been submitted to the journal *IEEE Transactions on Industrial Electronics*, under the title: “Generative Adversarial Networks for Acoustic Directivity Modelling in Parametric Arrays.” The manuscript is currently under review.

I used Grammarly only for proofreading-type suggestions. No academic content (ideas, analysis, equations, figures or code) was generated or rewritten. This complies with the Policy on Transparency of Authorship in PGR Programmes.

List of Publications

[1] **Mei, S**., Hu, Y., Alkahtani, M., & Esfahani, M. N. (2025). A Dual Convolution Model for Grating Lobe Prediction in Directivity of Parametric Array Loudspeakers. IEEE Access.

[2] **Mei, S**., Xu, H., Hu, Y., Alkahtani, M., & Wang, Y. (2022, November). The Parametric Array Speaker: A Review. In International Joint Conference on Energy, Electrical and Power Engineering (pp. 1254-1271). Singapore: Springer Nature Singapore.

[3] **Mei, S**., Hu, Y., Xu, H., & Wen, H. (2022). The class D audio power amplifier: A review. Electronics, 11(19), 3244.

[4] **S. Mei,** Y. Hu, and M. N. Esfahani, "Adaptive optimisation of non-uniform phased array speakers using particle swarm optimisation for enhanced directivity control," submitted to IEEE/ACM Transactions on Audio, Speech, and Language Processing, under review.

[5] **S. Mei,** Y. Hu, M. N. Esfahani, X. Zhang "Generative adversarial networks for acoustic directivity modelling in parametric arrays," submitted to *IEEE Transactions on Industrial Electronics*, under review.

Nomenclature

A_c	Carry wave Amplitude
A_{n}	Transducer Element Amplitude
C	Computational complexity
C_a	Quadratic-source constant
D	Space Dimensionality
$D(X)$	Discriminator output
$D_j(\varphi)$	Directivity of the array
D_{SPL}	SPL Directivity
$D_p(\varphi,\varphi_v)$	Directivity of the phased array
$D_w(\varphi)$	Westervelt's directivity function
$\widetilde{D_P}(\varphi)$	Dual convolution directivity
$D_j^{(3D)}(\Omega)$	Primary 3D directivity (element pattern × array factor)
$\bar{D}_P^{(3D)}(\theta,\varphi)$	3D dual-convolution total directivity
$D_{\mathrm{w}}^{(3D)}(\gamma)$	3D Westervelt directivity function
$E(X;\theta_e)$	Encoder function in GAN
$G(x)$	Generator function in GAN
$G_a(\mathbf{r})$	DF free-space Green's function
H_0	First-order Hankel function
$I_{\min}, I_{\max}$	Minimum and maximum intensity values in the image

K_a	Audio wavenumber
L_p	Sound Pressure Level (SPL)
$L_{\text{Main}}, L_{\text{Side}}^{(k)}, L_{\text{Min}}$	Main lobe/ Side lobe/ Minimum Main lobe
L_D	Discriminator loss function in GAN
L_G	Generator loss in GAN
L_{acoustic}	Physics-informed loss for acoustic consistency
L_{adv}	Adversarial loss for GAN
L_{L1}	L1 loss function for GAN training
N	Population Number
$P(\varphi)$	Summation of phased array components
$P(\theta_k)$	Far-field polar response of a uniform linear array
P_{ref}	Reference sound pressure
Q_S	Volume flow rate per unit length
S	Monopole amplitude
$S_a(\mathbf{r}_v)$	Quadratic source term
T_0	Standard air temperature
V_{IN}	Input signal energy
$V_{\text{h1}}, V_{\text{h2}}, \ldots, V_{\text{h}n}$	Harmonic voltage levels
W_d, b_d	Decoder weights and biases
W_e, b_e	Encoder weights and biases
$W_p(\varphi_v)$	Weight matrix for phase shift
X_{input}	Input sample (either real or generated)

Z_{latent}	Latent representation in GAN encoder
Z_l	Feature map at layer l
Z_{skip}	Skip connection in U-Net
c_0	linear speed of sound in air
c_c	Complex speed of sound
c_p	Cognitive Coefficient (Personal Learning)
c_s	Social Coefficient (Social Learning)
d, d_{limit}	Gap between transducers, Gap Limit
f_a ,f_c	Audio frequency/Carrier frequency (Hz)
f_{min} ,f_{max}	Minimum/ Maximum operational frequency(Hz)
$f_{r,N}$,$f_{r,O}$	Nitrogen/Oxygen relaxation frequency
$g(t)$, $g^{\wedge}(t)$	Input audio signal/ Hilbert transform of the input signal
g_{best} ,g_{besti}	Global Best Position, History Best Position
h	Absolute humidity
h_r	Relative humidity
k, k_i	Wavenumber
k_a	Audio (difference-frequency) wavenumber
k_{eq}^2	Effective wavenumber
m	Modulation index (depth)
p_a	Air pressure in the propagation environment
p_b	Nonlinear correction to pressure
p_d	Dipole domain source

$p_j(\rho_v)$	Acoustic pressure of ultrasound at the virtual source
p_r	Baseline ambient air pressure
p_t	Total acoustic pressure
p_{sat}	Saturated water vapour pressure
p_{best}	Personal Best Position
$q_u(\rho_v)$	Source density of the audio sound
q_d	Dipole domain source
r	Random number
$\hat{\mathbf{r}}$	Observation direction unit vector
$s(t)$	Modulated ultrasonic output signal
w_n	Complex weight for array element n
x_{ij}	Particle position in PSO
α	Sound attenuation coefficient
α_a	Difference frequency attenuation coefficient
a_p	Aperture Size
β	Nonlinear coefficient in air
δ	Sound diffusivity
$\delta(z\text{-}z_0)$	2D Delta function introducing the source
Δk	Wavenumber difference
λ	Wave Length
λ_1,λ_2	Weighting factors for physics-based losses
ρ_v	Virtual source point

ρ_0	The ambience air density
ω_c	Carrier angular frequency $\omega_c=2\pi f_c$
ω_a	Difference audio angular frequency
φ	Azimuthal angle
φ_0	Steering/reference azimuth angle
ψ	Optimal Steering Angle
$\sigma(\cdot)$	Activation function
τ	Delay between transducers
θ	Steering Angle
Ω_v	Virtual-source direction on the unit sphere
γ	Angle between observation and virtual-source directions
Aperture	Array aperture size
Noise	Total non-harmonic energy
Spacing	Transducer spacing
THD +N	Total Harmonic Distortion + Noise
ger	Maximum Iteration Number

Chapter 1

Introduction

1.1 Motivation

Precise control of sound direction is increasingly important in modern audio applications such as spatial sound rendering, personal audio systems, and targeted communication. To meet this demand, technologies capable of producing highly focused acoustic fields have gained considerable attention. Parametric Array Loudspeakers (PALs) stand out for their ability to generate narrow sound beams by leveraging the nonlinear properties of air. Unlike conventional loudspeakers that emit sound uniformly in all directions, PALs encode audible signals onto high-frequency ultrasonic carriers. As these waves propagate, they undergo self-demodulation due to nonlinear interactions in air, resulting in the formation of tightly focused audible sound fields [1]. This unique property enables precise sound delivery in public spaces, spatial audio systems, targeted communication, and active noise control (ANC) environments [2], [3]. However, achieving accurate and efficient prediction of PAL far-field directivity remains an open challenge due to the complexity of nonlinear sound propagation, beam attenuation, and the impact of transducer configuration.

To address this challenge, researchers have proposed three primary classes of modelling approaches. First, analytical methods based on the Westervelt equation, and its derivatives offer insight into nonlinear acoustic behaviour through simplified, closed-form solutions. These models are useful for theoretical analysis but often oversimplify boundary conditions and struggle to capture beam distortion or side-lobe behaviour in realistic setups [4]. Second, numerical methods, particularly Finite element method (FEM) simulations, provide high-fidelity modelling of acoustic propagation in nonlinear media. These methods are well-suited for capturing complex interactions and boundary geometries, but their computational cost severely limits their use in optimisation tasks or real-time applications [5]. Third, data-driven

approaches, including recent developments in deep learning, offer promising alternatives by learning complex mappings from simulation or experimental data. Among these, Generative adversarial networks (GANs) stand out for their ability to generate accurate, high-resolution outputs from limited inputs while preserving physical consistency [6], [7].

This thesis proposes an integrated modelling and optimisation framework for PALs by combining analytical theories, numerical simulations, and deep learning techniques. The primary motivation is to bridge the gap between accurate but computationally intensive numerical methods and fast, generalizable predictive models, while also addressing practical challenges such as beam steering and grating lobe suppression. The proposed framework aims to enhance the accuracy, efficiency, and scalability of far-field directivity prediction and array configuration for real-world PAL applications. The specific goals of this study are structured according to the major contributions of each chapter, as follows:

i. **To provide a structured review of acoustic technologies, the characteristics of PALs, and existing directivity modelling techniques.** By organizing existing methods into analytical, numerical, and data-driven approaches, the analysis reveals key limitations in current models, achieving real-time computation, and supporting steerable array optimisation. These insights help define the scope of the modelling challenges addressed in subsequent chapters.
ii. **To improve the accuracy of analytical directivity predictions through a dual convolution-based modelling strategy that overcomes the limitations of traditional approaches.** Conventional analytical models, though efficient, often overlook nonlinear propagation and struggle to capture grating lobes and edge effects. By integrating both spatial impulse response and transducer layout geometry, the dual convolution approach provides more accurate predictions of far-field behaviours. In our benchmarks against measured directivity, root-mean-square error (RMSE) decreases from 45.23 dB (convolution) to 15.51 dB (dual convolution), and main-lobe RMSE from 24.59 dB to 6.84 dB, while retaining lightweight computation, offering an interpretable tool for early-stage array analysis.
iii. **To optimize steerable PAL array configurations using metaheuristic algorithms.** Steering acoustic beams in PAL systems involves trade-offs between beam sharpness, side-lobe suppression, and physical layout. We implement a Particle swarm optimisation (PSO) based optimisation scheme is implemented to identify optimal

transducer placements under varying steering angles. PSO-optimised, non-uniform PAL layouts were validated by FEM and 2.0 m measurements (2.5° grid). Compared with uniform arrays, they consistently strengthened the main beam and suppressed grating lobes in the mid-to-high steering range (60°–75°), achieving side lobe reductions of up to 5 dB while maintaining acceptable beamwidth; trends were consistent across simulation and experiment.

iv. **To develop a GAN-based far-field prediction framework trained on high-fidelity FEM data to accelerate the simulation process**. Since FEM is computationally intensive, a deep generative model is proposed to approximate the complex nonlinear wave propagation patterns captured by simulation. A deep generative model is developed to learn the mapping from near-field or configuration data to far-field distributions. The model incorporates physics-informed loss functions to improve generalization. We assess accuracy via Mean absolute error (MAE) and RMSE against FEM/measurement, reporting MAE of 0.0105 and RMSE of 0.0171. The model delivers high-fidelity predictions with a substantial per-pattern runtime speedup relative to FEM, enabling scalable exploration of array configurations.

1.2 Dissertation Outline

This dissertation is organized into six chapters, each progressively developing the research from theoretical modelling to experimental validation and deep learning-based prediction. The structure is as follows:

Chapter 1 introduces the motivation and research background for the study, emphasizing the significance of accurately predicting the far-field directivity of PALs. The key research challenges are outlined, along with the objectives of the dissertation.

Chapter 2 presents a comprehensive literature review on PALs, focusing on their working principles, signal modulation schemes, transducer structures, and beamforming capabilities. The design of transducer arrays, including miniaturized PMUTs and non-uniform configurations, is discussed for their role in shaping directional performance. The chapter also classifies and evaluates three major modelling approaches for PAL directivity prediction: analytical models based on nonlinear acoustic theory, numerical simulations using FEM, and recent deep learning approaches such as GANs. Through this review, key challenges in

achieving accurate, efficient, and scalable modelling are identified, forming the basis for the subsequent research.

Chapter 3 presents the theoretical modelling of the far-field directivity of PALs, deriving the Westervelt-based convolution framework in 2D and its 3D counterpart and simulation of both primary and secondary acoustic fields. A Dual Convolution Model is proposed to address limitations in existing analytical methods. We benchmark the dual-convolution against the conventional convolution model, showing marked error reductions on measurement-referenced curves and improved grating-lobe fidelity. However, the study also identifies challenges in predicting far-field variations for complex or non-uniform arrays, motivating the need for more flexible modelling approaches.

Chapter 4 integrates FEM simulations with PSO to design sparse transducer arrays. We formulate a principal-plane 2D FEM for routine analysis and its 3D counterpart for scoping and asymmetric scenarios, clarifying when each is required. FEM accurately models nonlinear acoustic propagation, while PSO efficiently searches for optimal array configurations under steering constraints. To validate the simulation outcomes, an experimental platform is constructed, including a PAL prototype, microphone array, and data acquisition system. Measurement results of the sound pressure fields are compared against FEM predictions, confirming the effectiveness of the optimized arrays and revealing the trade-offs between accuracy and computational cost in practical implementation.

Chapter 5 introduces a GAN-based learning framework trained on FEM-generated datasets. The architecture and training methodology of the GAN model are detailed, with an emphasis on incorporating physics-informed loss functions to ensure physical plausibility. Techniques such as data augmentation, domain adaptation, and structured learning are employed to enhance generalization. The model is evaluated on its ability to predict far-field directivity patterns under various PAL configurations, demonstrating both accuracy and efficiency improvements over conventional simulation methods.

Chapter 6 summarizes the main contributions of this work, including a dual convolution model for grating lobe prediction, PSO-based array optimisation, and a GAN-based method for fast far-field directivity prediction from sparse data. These methods are validated through simulations and experiments. Future work will focus on extending the models to 3D, improving data acquisition, and integrating physics-informed constraints into deep learning frameworks.

Chapter 2

Review of Parametric Array Loudspeaker and Directivity Prediction

2.1 Introduction

Parametric array loudspeakers (PALs) are drawing increasing interest. They deliver tightly focused sound in open air by using ultrasound and nonlinear demodulation. This capability supports uses from immersive installations and museums to public address and wayfinding, where precise coverage and low spill matter. [1], [8], [9].

The ability to control the directivity of PALs is central to their performance in real-world deployments[10]. However, accurately predicting far-field sound remains challenging due to several interrelated factors. These include the nonlinear nature of acoustic wave propagation, the influence of transducer arrangement and spacing, the selection of modulation schemes, and the potential emergence of grating lobes in sparse or irregular arrays. In addition to hardware considerations, digital signal processing (DSP) also plays a key role in shaping and steering the emitted wavefronts, further complicating the modelling of directivity [11].

Researchers use three modelling methods to manage these effects. Analytical models, such as Westervelt-based convolution, give clear insight into nonlinear propagation. They are fast and easy to analyse, but their simplifying assumptions reduce accuracy in complex scenes. Numerical models, especially FEM, raise fidelity and capture boundaries and coupling well. The trade-off is heavy computation, which limits parameter sweeps and real-time use. Data-driven models offer a third path. Recent GAN-based frameworks learn rich mappings and can predict fast while keeping strong accuracy once trained.

This chapter provides a comprehensive review of the key technologies and modelling methods associated with PAL directivity. It begins with an overview of the physical principles, transducer architectures, and digital modulation strategies that define PAL operation. And then examines various application scenarios that highlight the importance of accurate directivity

control. Finally, the chapter evaluates existing modelling techniques, including analytical and numerical approaches, deep learning methods, and hybrid frameworks, to identify opportunities for improved prediction and control of PAL sound fields in complex environments.

2.2 Background of Parametric Speakers

PALs represent an advanced directional sound technology that utilizes ultrasonic waves to project audible sound in a highly focused manner. Unlike traditional loudspeakers that emit sound omnidirectionally, PALs modulate audio signals onto high-frequency ultrasonic carriers. As these carriers propagate through air, nonlinear interactions cause self-demodulation, reconstructing the audible signal only within the target area [2]. This high directivity minimizes acoustic spillover, making PALs well-suited for applications requiring precise and localized sound delivery.

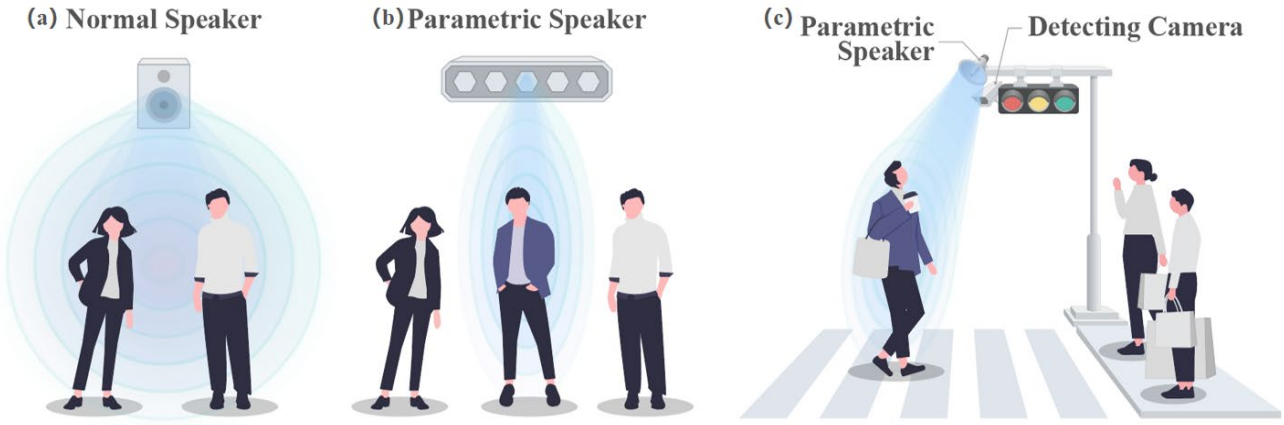


Fig. 2-1. The sound propagation characteristic of parametric speaker (a) loudspeaker with wide dispersion; (b) parametric loudspeaker with a narrow, highly directional beam; (c) targeted crosswalk announcement using a parametric loudspeaker triggered by a detecting camera [11].

As shown in Fig. 2-1, the illustration compares conventional and PALs and highlights a practical application scenario. In panel (a), a traditional loudspeaker radiates sound uniformly in all directions, resulting in uncontrolled dispersion. Panel (b) depicts a PALs generating a highly directional beam that enables precise sound localization. Panel (c) illustrates a pedestrian alert system in which a camera triggers the PAL to deliver personalized audio cues

to individuals at a crosswalk. This targeted sound projection enhances intelligibility while minimizing ambient noise, making it particularly suitable for urban environments [11].

2.2.1 Application for Parametric Speakers:

A major advantage of parametric speakers is their ability to create focused audio field, which means sending sound to specific people or areas while keeping the surroundings quiet. This highly focused sound projection has led to numerous applications across various domains.

a) Healthcare and Assistive Technologies

PALs are increasingly used in healthcare. They deliver focused sound to one listener without affecting nearby people. This feature is promising for music therapy in dementia care, where speakers or headphones can be disruptive or impractical. When the system directs sound to the target listener, it can reduce agitation and wandering. It also helps keep a calm environment for others. A study by Yuko and colleagues showed that parametric audio reduced dementia-related symptoms in a residential care facility [12]. Typical installations place the emitters near beds or chairs. They can also link to presence detection to trigger simple sound cues. However, challenges remain, including discomfort at high frequencies and interference in complex sound fields [12].

b) Transportation and Safety Systems

In urban transportation systems, PALs have been investigated for their ability to deliver targeted warnings without contributing to ambient noise. A typical application involves pedestrian alert systems at crosswalks, where cameras or sensors detect individuals and trigger a narrow sound beam directed only at those in potential danger. Compared to traditional speakers or broadcast alarms, PALs can provide discrete yet effective communication**,** improving both safety and environmental noise control. Studies have also explored their use in bus-pedestrian collision prevention systems**,** where directional alerts reduce reaction time without disturbing surrounding traffic or pedestrians. These implementations often involve integration with vision systems and require careful alignment to ensure accurate sound targeting, especially in dynamic outdoor environments [13], [14].

c) Scientific and Industrial Applications

PALs have been increasingly adopted in scientific and industrial domains due to their ability to generate precisely controlled acoustic beams in the air. In atmospheric research, they play a key role in Radio Acoustic Sounding Systems (RASS)**,** where modulated ultrasonic signals are transmitted vertically to probe air temperature and wind velocity through acoustic wave reflections. Compared to conventional sources, PALs offer better beam collimation and frequency stability, resulting in higher measurement accuracy under varying environmental conditions [15].

In industry, PALs are added to active noise control (ANC) to quiet local zones, such as workstations, control panels, or passenger areas. Traditional ANC often needs many loudspeakers or large barriers. By contrast, PALs project phase-inverted sound along a narrow path. They achieve cancellation without affecting nearby areas. This enables open-space quiet zones and improves comfort and speech clarity [16].

PALs are also explored for industrial monitoring and ultrasonic testing where precise spatial excitation is required. However, challenges remain. Hard surfaces cause reflections. Beam steering can be limited. Power is often constrained in large or reverberant spaces. Ongoing research aims to improve beam shaping, source modulation, and array control to raise reliability in complex acoustic conditions [17].

d) Targeted Audio and Immersive Experiences

Given their high directivity and sound beamforming capabilities, parametric speakers have significant potential in spatial audio, immersive sound experiences, and intelligent communication systems. They have been applied in museum and exhibition audio guides, where highly localized sound projection allows visitors to receive audio information without disrupting others nearby. The studies [18] have explored their use in acoustic measurement setups, where their precision and directionality contribute to improved sound field control and measurement accuracy.

2.2.2 Fundamentals and System Architectures of Parametric Speaker

Parametric array loudspeakers operate based on principles of nonlinear acoustic propagation, using ultrasonic carrier waves, typically in the range of 20 kHz to 100 kHz, to transmit audio

signals. These high-frequency waves are modulated with audible content and emitted into the air, where the medium's inherent nonlinearity triggers a self-demodulation process. This results in the reconstruction of the audible signal at a distance, confined to a narrow propagation region. Due to the short wavelength of ultrasonic waves, PALs achieve minimal diffraction and high directionality, allowing precise spatial targeting of sound energy [19], [20].

In this study, a 40 kHz carrier wave is used, chosen for both its acoustic advantages and engineering practicality. First, it exceeds the upper limit of human hearing (20 kHz), ensuring inaudibility. Second, 40 kHz is a standard frequency in airborne ultrasonics, with numerous cost-effective and readily available transducers such as the Murata MA40S4 used in our setup which designed for optimal performance at this frequency. Third, while higher ultrasonic frequencies suffer from rapid air attenuation, 40 kHz offers an effective trade-off: it is sufficiently high to trigger nonlinear self-demodulation while still propagating well through air without excessive loss [21].

Unlike conventional loudspeakers that rely on diaphragm vibration, parametric speakers emit modulated ultrasound that demodulates during propagation, producing audible sound solely along a narrow beam path. The short wavelength of ultrasonic waves minimizes diffraction and enhances directivity, a phenomenon explained by Huygens' principle [21]. This capability allows sound to reach a specific spot or a single listener while nearby areas remain unaffected. It suits museum guidance, public announcements, active noise control (ANC), and spatial audio in smart environments.

However, PALs have technical limits. Effective self-demodulation often needs sound pressure level (SPLs) exceeding 110 Decibels (dB), which raises power demand. The nonlinear process also creates harmonic distortion and reduces low-frequency output. In addition, air absorption at ultrasonic frequencies lowers transmission efficiency. To address these issues, researchers study advanced modulation, improved transducer design, and high-efficiency power amplification. Recent work on hybrid modulation, adaptive beamforming, and AI-enhanced DSP shows promise for better performance and energy efficiency [22], [23].

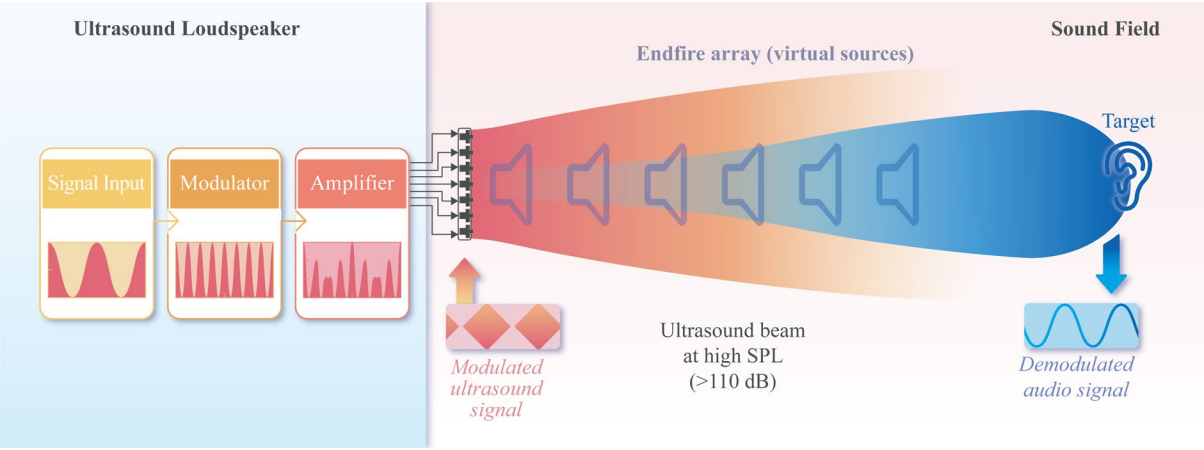


Fig. 2-2. The infrastructure of the parametric speaker and the demodulation process in the air.

As shown in Fig. 2-2, a typical parametric speaker system consists of several key components: audio signal input, modulation stage, power amplification, ultrasonic transducer array, and the demodulation region where audible sound is perceived. The focused beam undergoes nonlinear self-demodulation during propagation [25], effectively forming a virtual sound source at the target. Upon interacting with a listener or obstacle, the ultrasonic and audible components separate, ensuring only the desired listener perceives the sound.

- **DSP Unit** – Modulates the input audio signal onto an ultrasonic carrier wave. The most used modulation techniques include Amplitude Modulation (AM) and Frequency Modulation (FM). These modulation techniques determine how efficiently the sound is encoded into the ultrasonic wave for effective transmission.

- **Power Amplifier** – Amplifies the modulated signal to a sufficiently high level to drive the transducers while maintaining minimal distortion. Given the high frequencies involved, power amplifiers used in parametric speakers must exhibit wide bandwidth and high efficiency to prevent excessive energy loss.

- **Transducer Array** – Converts the amplified electrical signal into an ultrasonic beam with high directivity. This array typically consists of multiple piezoelectric transducers arranged to produce a narrow sound field, allowing for focused sound projection with minimal dispersion.

2.2.3 Digital Modulation Methods for Parametric Speakers

Modulation is a very important process to parametric speakers. It sets how the ultrasonic carrier encodes the audio and how the air later reconstructs it. Conventional loudspeakers generate audio directly. Parametric speakers do not. They rely on nonlinear effects in air to self-demodulate the transmitted signal. This modulation pathway produces a highly directional sound field. It is essential for applications that need localised audio delivery.

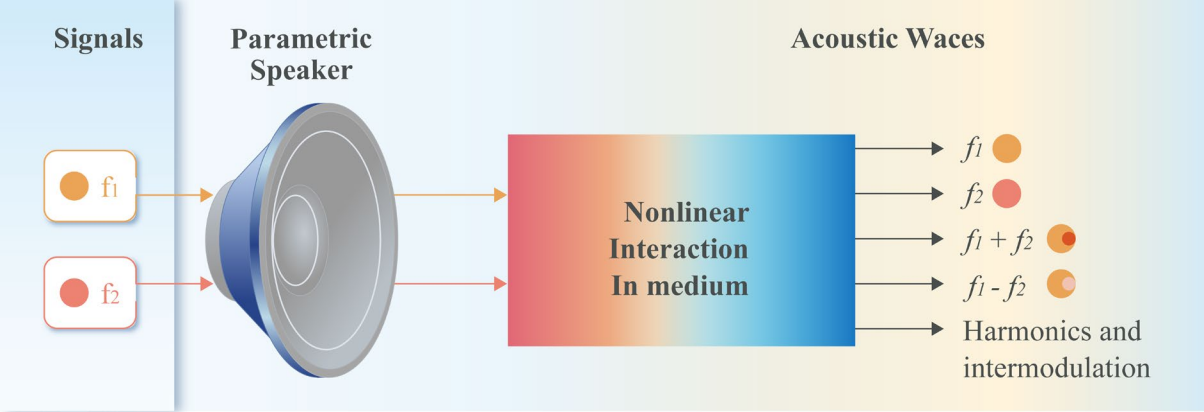


Fig. 2-3. Nonlinear interaction process in air for signal generation in PALs.

As illustrated in Fig. 2-3, a parametric speaker is typically driven by two closely spaced ultrasonic signals, denoted as f_1 and f_2. Due to the nonlinear interaction within the air medium, these signals generate additional spectral components, including the sum frequency (f_1+f_2), the difference frequency (f_1-f_2) which is also be called DF, and various higher-order harmonics such as $2f_1$,$2f_2$, and intermodulation products. Among these components, the higher frequency harmonics experience rapid attenuation in air, diminishing quickly as the distance from the speaker increases. In contrast, the DF (f_1-f_2), which falls within the audible range, is comparatively less absorbed by air, allowing it to propagate further while inheriting the spatial characteristics of the ultrasonic waves. This process enables parametric speakers to achieve an extremely narrow beam of audible sound, maintaining high directivity and minimal sound dispersion even over long distances[8], [25].

This section will further discuss the different modulation techniques used in parametric speakers, including AM, Single-Sideband Modulation (SSB), and DSB methods, which enhance the efficiency and precision of directional sound reproduction.

a) Double Sideband (DSB) amplitude modulation

DSB Modulation is one of the earliest and most widely used modulation schemes in parametric speaker systems, initially introduced in 1983 [26] which is also used in 3.3 as the modulation method. In this approach, an audio signal is embedded into a high-frequency carrier via amplitude modulation, producing a modulated waveform that undergoes nonlinear self-demodulation in air. The resulting audible signal is formed through the interaction of primary ultrasonic components, governed by the modulation envelope. A typical DSB-AM waveform is defined as:

$$s(t)=A_c[1+mg(t)]\cos\left(2\pi f_c t\right) \tag{2-1}$$

where m is the modulation index (depth), g(t) is the (normalized) audio message, f_c is the carrier frequency, and A_c is the carrier amplitude.

As illustrated in Fig. 2-4, the DSB-AM process consists of:

- Scaling the audio g(t) by m to set the modulation depth;
- Adding a DC term to form the envelope $1+mg(t)$ (keeps the envelope non-negative when $m \leq 1$);
- Multiplying by the carrier $A_c\cos\left(2\pi f_c t\right)$ to obtain $s(t)$.

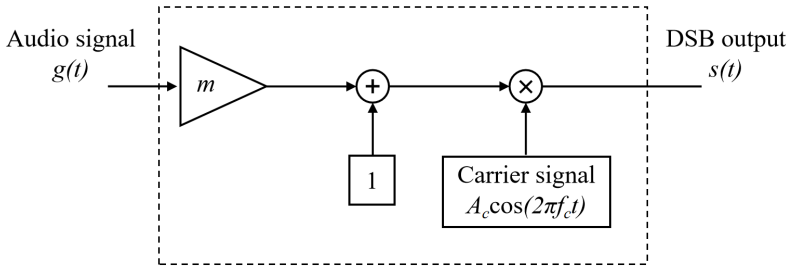


Fig. 2-4. Block diagram of DSB-AM with carrier signal generation.

While DSB-AM is simple to implement, it presents a trade-off between demodulation efficiency and harmonic distortion. As shown in [27], increasing the modulation index improves signal strength but also leads to a substantial rise in total harmonic distortion (THD), reaching up to 80% when m = 1. This high distortion compromises sound quality and limits practical deployment. The limits of DSB-AM motivate alternative schemes. Single-sideband

(SSB) and recursive amplitude modulation (RAM) are two common options. They aim to raise fidelity and spectral efficiency. They also reduce total harmonic distortion (THD). As a result, SSB and RAM are often better choices for high-performance parametric speaker designs.

b) Modified amplitude modulation (MAM)

To improve the efficiency of parametric speaker systems, [28], [29] propose a MAM method, which combines Amplitude Modulation (AM) and Square-Root Amplitude Modulation (SRAM) based on orthogonal AM principles. The MAM method is designed to optimize the trade-off between audio loudness and distortion, which is a critical consideration when choosing a modulation technique and modulation index. The block diagram of the MAM method is illustrated in Fig. 2-5.

The MAM modulation signal can be expressed as:

$$s(t)=(mg(t)+1)\cos(\omega_c t)+\sqrt{1-(mg(t))^2}\cos(\omega_c t) \tag{2-2}$$

Where $s(t)$ is the modulated output signal, which is transmitted as an ultrasonic wave. $g(t)$ present input audio signal, containing the low-frequency audible content. And m controlling the depth of modulation:

- Higher m increases loudness but also raises distortion.
- Lower m reduces distortion but weakens the output.

This method introduces a square-foot function to the modulation envelope, which helps reduce nonlinear distortion by controlling the depth of modulation. As a result, the modulated signal maintains a more linear response, leading to improved demodulation efficiency and reduced harmonic artifacts.

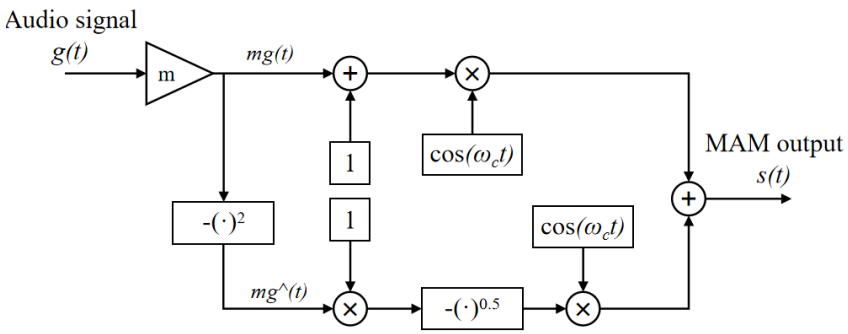


Fig. 2-5. Block diagram of the MAM method.

A major challenge in modulation design is balancing audio loudness and signal fidelity. A higher modulation index increases demodulated signal strength, making the output more audible, but also introduces higher harmonic distortion [30]. To minimize distortion, one approach is the square-rooting method, which modifies the amplitude envelope to reduce the influence of high-order nonlinear components. However, this approach requires wide bandwidth ultrasonic transducers to handle the modified signal efficiently [31].

While MAM improves modulation linearity, it adds complexity to the system design, especially in terms of hardware implementation and transducer response. Despite this, MAM remains a viable alternative to conventional AM techniques for parametric speaker systems where low distortion and high sound clarity are prioritized.

c) Single-Sideband (SSB)

An alternative strategy to reduce distortion and improve spectral efficiency is the Single-Sideband with Carrier (SSB-WC) system, as illustrated in Fig. 2-6. Compared to conventional AM, SSB modulation utilizes electrical power and bandwidth more efficiently, making it a preferred choice for parametric speaker applications [32].

The SSB-WC signal equation is given by:

$$s(t)=m[g(t)\cos(\omega_c t)\pm \hat{g}(t)\sin(\omega_c t)]+\sin(\omega_c t) \qquad (2\text{-}3)$$

Where, $g(t)$ is the signal, $\hat{g}(t)$ represents the Hilbert transform of the input signal, and ω_c is the carrier angular frequency, the diagram of the SSB-WC has been shown in Fig. 2-6.

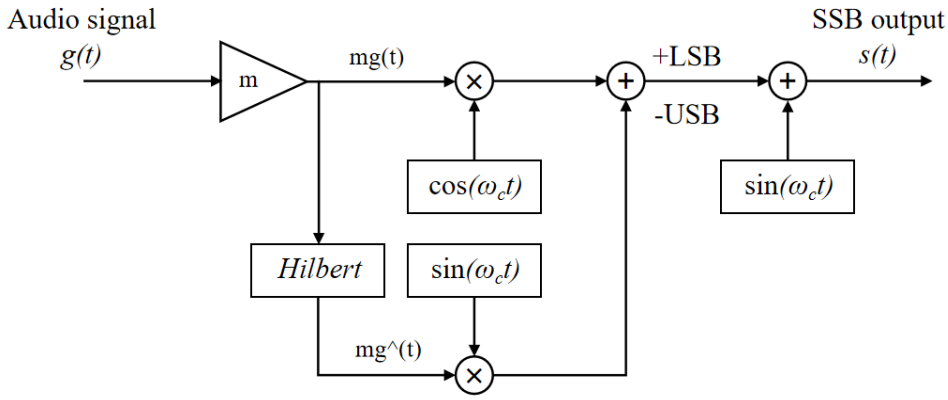


Fig. 2-6. Block diagram for single sideband modulation with carrier (SSB-WC).

In Fig. 2-6, the positive (+) and negative (-) symbols denote Lower Sideband (LSB) and Upper Sideband (USB) modulation, respectively. The Hilbert transform block in the diagram follows a transfer function $H(\omega)=-j\mathrm{sgn}(\omega)$, where sgn($\boldsymbol{\omega}$) is defined as:

$$\mathrm{sgn}(\omega)=\begin{cases}1, & \text{for } \omega>0\\ 0, & \text{for } \omega=0\\ -1, & \text{for } \omega<0\end{cases} \tag{2-4}$$

For a normalized signal $g(t)=\cos(\omega t)$, the Hilbert transform introduces a phase shift of $\pi/2$, such that $g^{\wedge}(t)=\sin(\omega t)$. Consequently, the output of the single-sideband modulator $\varphi_{SSB}(t)$ is:

$$\varphi_{SSB}(t)=0.5m\cos(\omega_c t)\pm 0.5m\sin(\omega t) \tag{2-5}$$

which simplifies to:

$$\varphi_{SSB}(t)=0.5m\cos[(\omega_c \mp \omega)t]\# \tag{2-6}$$

For a single-tone input signal $\cos(\omega t)$, the modulated Lower Sideband with Carrier (LSB-WC) signal is(2-7):

$$\varphi_{LSB\text{-}WC}(t)=0.5\{\cos(\omega_c t)+m\cos[(\omega_c-\omega)t]\} \tag{2-7}$$

Similarly, the modulated Upper Sideband with Carrier (USB-WC) signal is2-8:

$$\varphi_{USB\text{-}WC}(t)=0.5\{\cos(\omega_c t)+m\cos[(\omega_c+\omega)t]\} \tag{2-8}$$

As shown in Fig. 2-6, SSB-WC reduces redundant frequency components, thereby improving demodulation efficiency while maintaining a high-quality, low-distortion audio output. Compared to DSB-AM, SSB modulation minimizes spectral waste and reduces the risk of nonlinear distortion, making it a more effective choice for high-performance parametric speakers.

2.3 Parametric Speaker Design and Transducer Array Configuration

While the unique directivity and low-spillover characteristics of PALs have enabled diverse applications, their practical effectiveness relies heavily on the physical design of the transducers and the configuration of the acoustic array**.** Unlike conventional audio systems, PALs must generate and control ultrasonic carriers with high precision, requiring specialized transducer materials, structures, and driving methods [33], [34]**.**

Equally important is the arrangement of transducers within an array, which directly affects beamwidth, grating lobe formation, spatial resolution, and SPL. Choices regarding aperture size, spacing, array geometry, and element phasing have a considerable influence on the resulting sound field and its adaptability to different environments.

The following sections examine critical aspects of ultrasonic transducer design for PALs, including recent developments in miniaturized, as well as array configuration strategies. These physical design elements form the foundation for precise beamforming and high-fidelity sound projection, which are central to achieving accurate far-field directivity, one of the primary goals of this study.

2.3.1 Beamforming Capability

Acoustic beamforming is a spatial filtering technique that employs DSP to control the direction of sound propagation. In parametric speaker arrays, beamforming enhances directivity and reduces unwanted sound dispersion by manipulating the phase relationships among transducer elements. By applying carefully calibrated phase delays, the system induces constructive interference in a desired direction, an approach derived from phased array technology [35], [36]. As illustrated in Fig. 2-7, increasing the number of transducers and optimizing the delay profiles sharpens the main lobe and enables electronic steering of the sound beam [37].

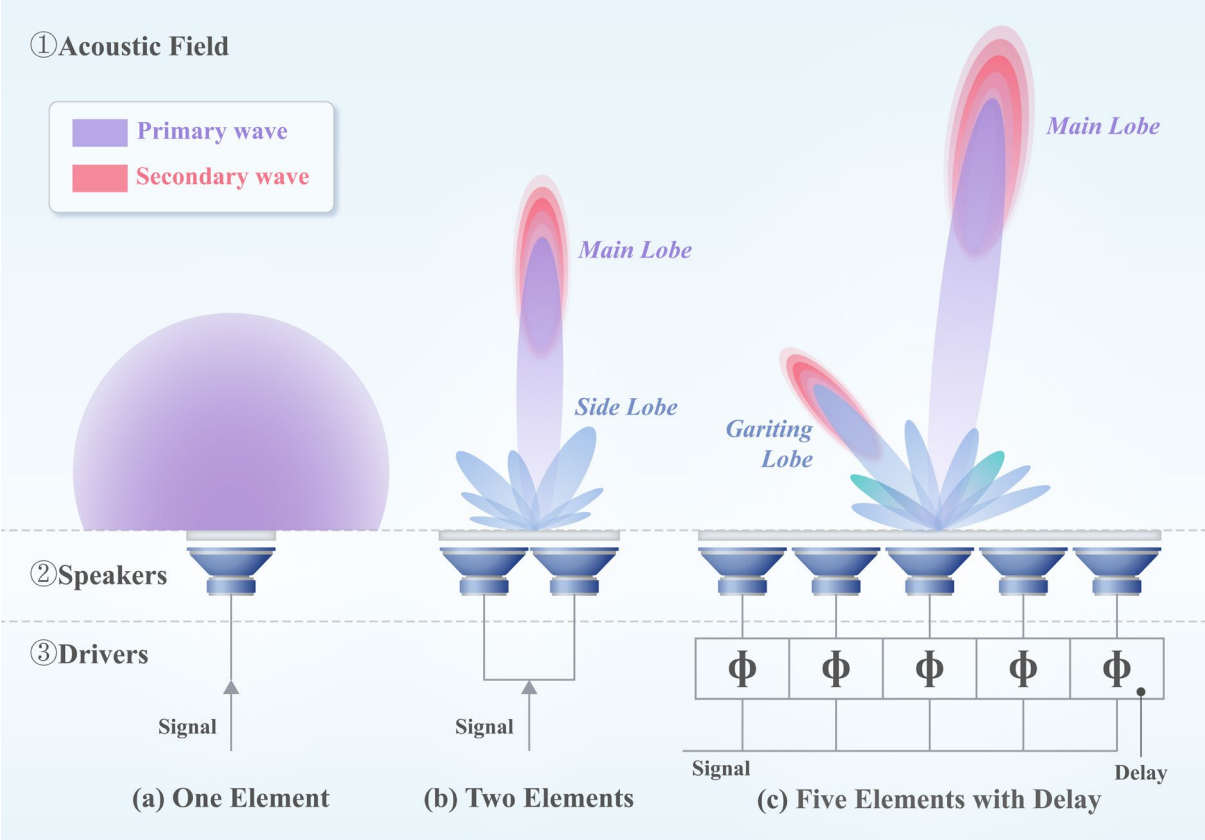


Fig. 2-7. Evolution of beamforming in parametric speaker arrays: (a) Single-element source with omnidirectional field, (b) Two-element array producing limited beamforming, and (c) Five-element array with delay-based phase control achieving focused directivity.

As shown in Fig. 2-8, beamforming in PAL systems is realized through a multi-transducer configuration with individually controlled phase delays. These delays generate a focused ultrasonic beam that undergoes nonlinear self-demodulation during propagation, forming a steerable audible sound field. Although the complete system may include Analog-to-digital Converters, modulators, and amplifiers, it is the integration of precise phase control with nonlinear acoustic effects that fundamentally enables directional sound projection.

Achieving beamforming in PAL systems requires driving each ultrasonic transducer with an independent amplifier, enabling individual phase control across the array. By exploiting constructive and destructive interference, a highly directional ultrasonic beam is formed, which self-demodulates into an audible signal through nonlinear interaction in air. This configuration functions as an end-fire virtual array, allowing precise targeting with minimal off-axis

dispersion. However, implementing such systems presents substantial challenges, particularly in maintaining synchronization across multiple amplifiers and preserving signal integrity [38].

Despite the challenges in implementation, advances in DSP have significantly enhanced the beamforming capabilities of PALs. Modern algorithms such as delay-sum, minimum variance distortion less response (MVDR), and adaptive filtering enable real-time steering of narrow acoustic beams with high spatial precision. Unlike traditional systems that rely on mechanical rotation or gimbals for beam redirection, parametric speakers with phase control can steer sound electronically by adjusting the delay and phase applied to each transducer. This eliminates the need for complex mechanical structures, leading to more compact, reliable, and rapidly reconfigurable audio systems.

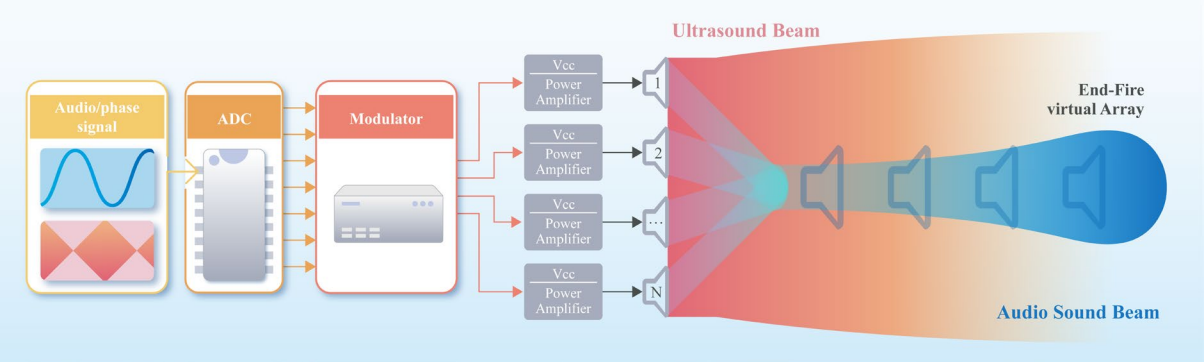


Fig. 2-8. Parametric array loudspeaker array with beamforming capability enabled by digital signal processing, using individually amplified ultrasonic transducers and phase-controlled modulation to generate a steerable audio beam.

2.3.2 Power Amplifier for Parametric Speaker

Parametric array loudspeaker systems, like conventional audio systems, can employ various power amplifier topologies, including Class A, B, AB, and D configurations. Analog amplifiers such as Class A and AB offer excellent signal fidelity but suffer from low power efficiency typically below 70% [39], [40]. In contrast, switching-mode amplifiers, particularly Class D, have emerged as a highly efficient alternative, achieving power conversion efficiencies exceeding 90% at high modulation indices [41]. As shown in Fig. 2-9, this efficiency advantage has led to the widespread adoption of Class D amplifiers in commercial parametric speaker designs, replacing many traditional analogue implementations [42], [43]. In this thesis, we employ a Class-D power stage to generate and deliver the ultrasonic drive: the DSB-AM carrier and audio envelope are composed in the digital domain and then amplified by a MOSFET H-

bridge to excite the array. Implementation details and measurement configurations are provided in Chapter 3 and in the experimental-setup sections of Chapters 3 and 4.

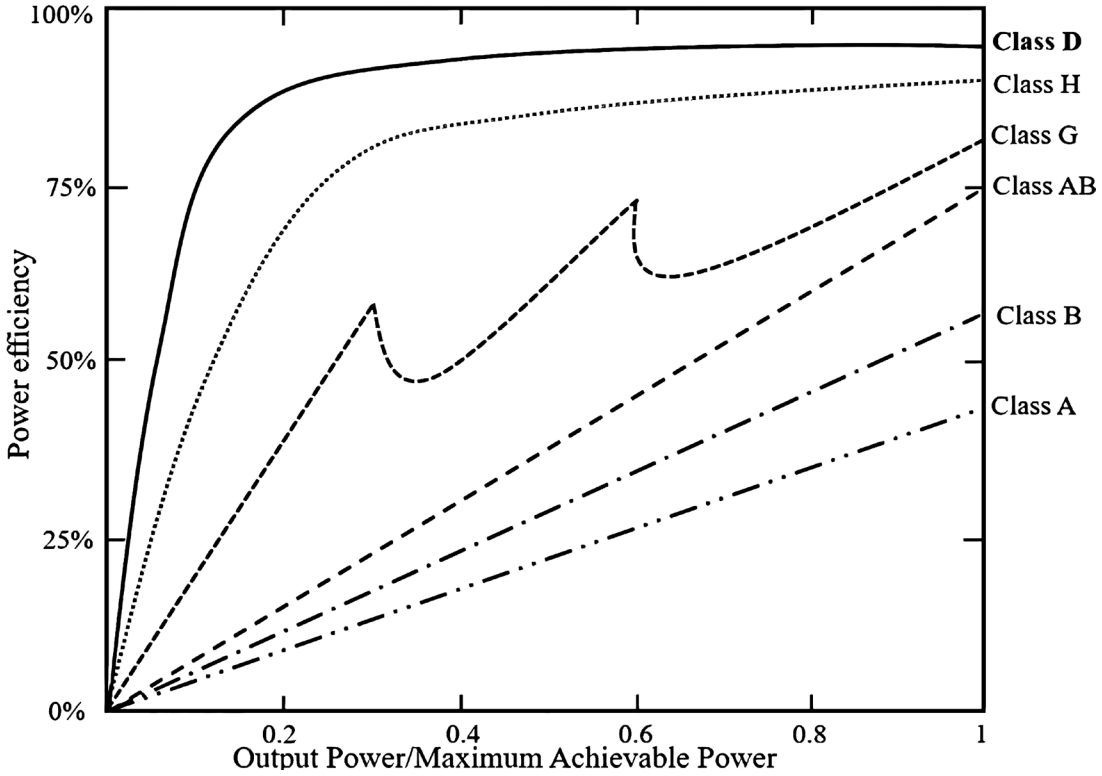


Fig. 2-9. Power efficiency comparison of amplifier classes used in audio and parametric array loudspeaker systems [32].

a) Analog Amplifier

Class A amplifiers are occasionally employed in parametric speaker systems due to their superior linearity and low distortion. These amplifiers operate continuously throughout the entire input signal cycle, including both positive and negative phases. As a result, they consume power even in the absence of an input signal, leading to poor power efficiency, ranging from 15% to 30% [44].

To address these inefficiencies, push-pull Class B amplifiers have been proposed as an alternative. Operating at a lower DC bias point than Class A amplifiers, Class B designs offer improved power efficiency and wider bandwidth. However, this comes at the cost of increased signal distortion due to crossover nonlinearity and reduced fidelity [45], [46].

To mitigate these issues, pre-linearized Class B amplifiers have been developed to enhance linearity by reducing voltage gain variation [47], [48]. Another option is the Class C amplifier,

a nonlinear configuration that offers high efficiency but suffers from reduced output power and limited sensitivity, making it less suitable for precision audio applications.

To overcome this limitation, a novel diode expander architecture has been proposed to improve Class C amplifier performance [49]. By increasing the input pulse amplitude applied to the power transistor, this structure enhances the output voltage and compensates for sinusoidal waveform degradation caused by limited conduction angles.

b) Digital Amplifier

Class D amplifiers adopt a switching topology that significantly reduces power dissipation compared to conventional linear amplifiers. As illustrated in Fig. 2-10, analogue audio signals are first sampled and converted into pulse-modulated waveforms, which are then used to drive output switching devices. The resulting high-frequency switching signals are filtered either by a dedicated low-pass filter or through the mechanical response of the speaker itself to reconstruct the audible signal.

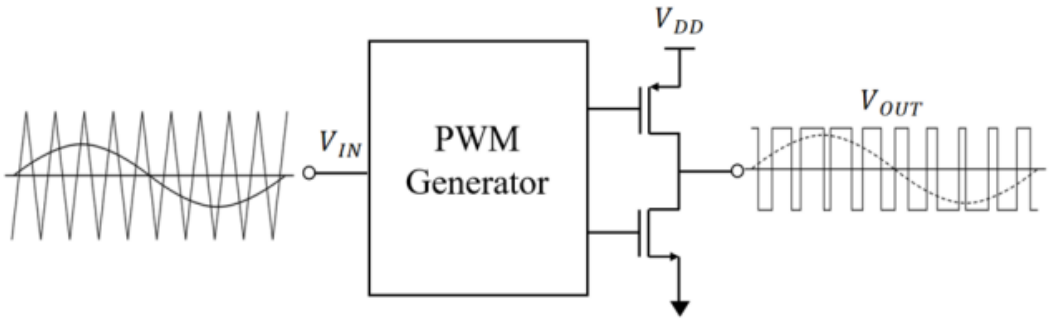


Fig. 2-10. Topology and signal waveform of a Class D amplifier using pulse-width modulation (PWM).

Depending on the modulation strategy employed, Class D amplifiers can be categorized into several types, including Pulse Width Modulation (PWM), Pulse Density Modulation (PDM), Delta-Sigma (Δ-Σ) Modulation, and Self-Oscillating schemes [50], [51] . Each technique offers distinct trade-offs in terms of spectral efficiency, distortion, and control complexity. In our implementation, a PWM-modulated audio signal is fed into the TC4427 MOSFET driver, which controls the switching of power transistors connected to a DC supply (V_{DD}). This setup forms a discrete Class D amplification stage, where the amplified switching output is delivered to the speaker. The acoustic signal is reconstructed either through a low-pass filter or the inherent mechanical response of the loudspeaker.

c) Semiconductor Materials of Amplifier

Wide-bandgap (WBG) semiconductors such as silicon carbide (SiC) and gallium nitride (GaN) have enabled the development of Class D amplifiers with improved distortion characteristics and wider bandwidths [52]. Compared to silicon (Si) devices, GaN transistors support higher operating voltages and exhibit lower leakage currents [53]. They also offer reduced on-resistance and conduction losses because of their higher electron saturation velocity, which is approximately 2.8 times greater than that of Si.

GaN devices also benefit from lower parasitic capacitance, allowing them to operate at switching frequencies in the megahertz range. Table 2-1 summarizes the key differences in electrical and thermal properties between GaN and Si devices. The wider bandgap, higher carrier mobility, and better thermal performance of GaN materials enable faster switching and lower conduction losses [54], [55].

Table 2-1. Comparison of Key Electrical and Thermal Parameters of Semiconductor Materials (Si, SiC, GaN).

Characteristic Parameters	Unit	Si	SiC	GaN
Energy gap	eV	1.1	3.26	3.49
Electron mobility	cm^2/vs	1500	700	2000
Saturated electron velocity	$10^7/s$	1	2	2.8
Electric breakdown field	MV/cm	0.4	2	3.3
Thermal conductivity	W/(cm K)	1.5	4.5	>1.5
Relative permittivity	εr	11.8	10	9

Experimental results confirm that GaN-based Class D amplifiers can achieve higher power output and lower THD compared to their silicon-based counterparts [56]. A comparative analysis of 60 V GaN and Si devices showed that GaN delivers significantly higher power density [57], [58]. In addition, GaN FETs are more compact, efficient, and cost-effective than traditional Si MOSFETs. One study demonstrated a compact eight-phase GaN Class D amplifier operating at 5 MHz for ultrasound cancer therapy, highlighting GaN's advantages in high-frequency acoustic applications [59].

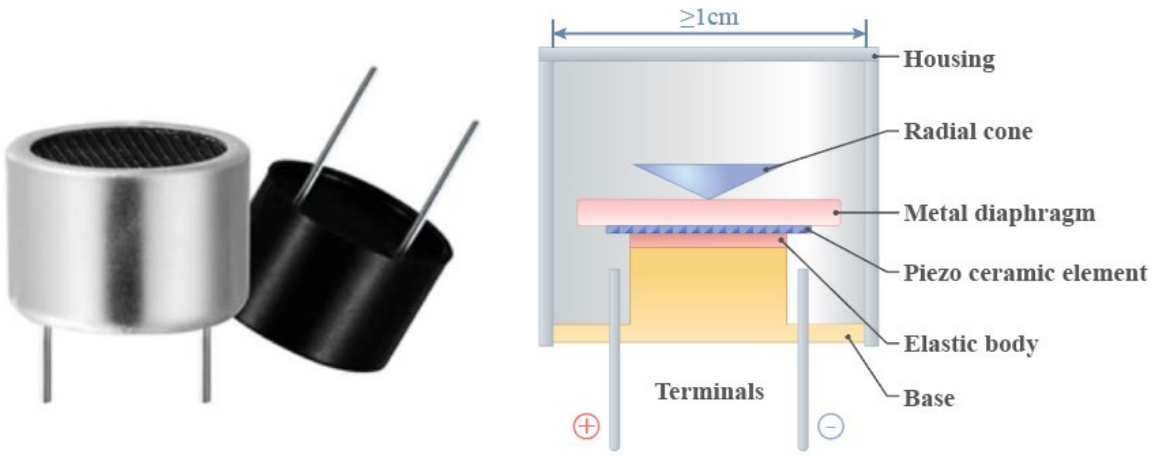
2.3.3 Design of Transducers

To keep a consistent and balanced frequency response, designers tune piezoelectric transducers for a broad resonance peak or for several distinct resonance modes. Recent work also targets miniaturised structures to fit PALs into compact electronic devices.

For example, a circular PAL in [60] uses a dual-diaphragm layout with two radial cones on a piezoelectric transducer. This layout raises acoustic output. It also uses an acoustic-crystal waveguide to produce highly directional ultrasound. In addition, piezoelectric micromachined ultrasonic transducers (PMUTs) are promising for compact PALs because they integrate well and use little power. [61].

Due to the impact of radiation structure size on radiation impedance, thin-plate PMUTs have been employed to enhance electro-acoustic efficiency. For example, [62] implemented a Complementary metal oxide semiconductor (CMOS)-compatible, air-coupled parametric array loudspeaker using dual-crystal aluminum nitride PMUTs, demonstrating a 200% improvement in output performance with a beamwidth that is narrower than 5 degrees. The PMUT array consisted of 246 transducer units integrated within a 13 mm × 13 mm chip. Additionally, to expand the operational bandwidth of the transducer array, out-of-phase driving was introduced, utilizing multiple radiators with different resonance frequencies [63].

These advancements have led to compact PAL designs that combine low power consumption with high-fidelity audio, making them suitable for portable devices such as laptops and smartphones. Fig. 2-11 shows the internal structure of a typical ultrasonic transducer, which includes a piezoelectric element that converts electrical signals into mechanical vibrations. These vibrations are transferred to a metal diaphragm and then radiated as ultrasonic waves. A radial cone shapes the beam direction, while the housing and elastic components provide structural support and vibration stability. This design enables efficient ultrasonic generation, essential for the precise sound control in PAL systems.



(a) (b)

Fig. 2-11. Package and internal structure of a piezoelectric ultrasonic transducer used in PALs. (a) External view; (b) cross-section with labelled components [64].

2.3.4 Configuration of Transducers

Transducer configuration is determined by the application and available installation space. Different array shapes produce varying frequency responses and directivity patterns, as the effective radiation length depends on array geometry. Higher element density generally narrows the beam but may reduce power efficiency [65].

Olszewski and Linhard analysed various parametric speaker array configurations and proposed a ring-shaped transducer arrangement to reduce beamwidth at lower frequencies, as illustrated in Fig. 2-12 (a). Furthermore, introducing phase deviations between transducer elements, as shown in Fig. 2-12 (b), enables additional acoustic effects, such as enhanced noise rejection, reduced overall noise levels, and 3D spatial sound reproduction. Additionally, attaching a horn structure to the parametric speaker, as depicted in Fig. 2-12 (c), improves beam collimation by further increasing directivity.

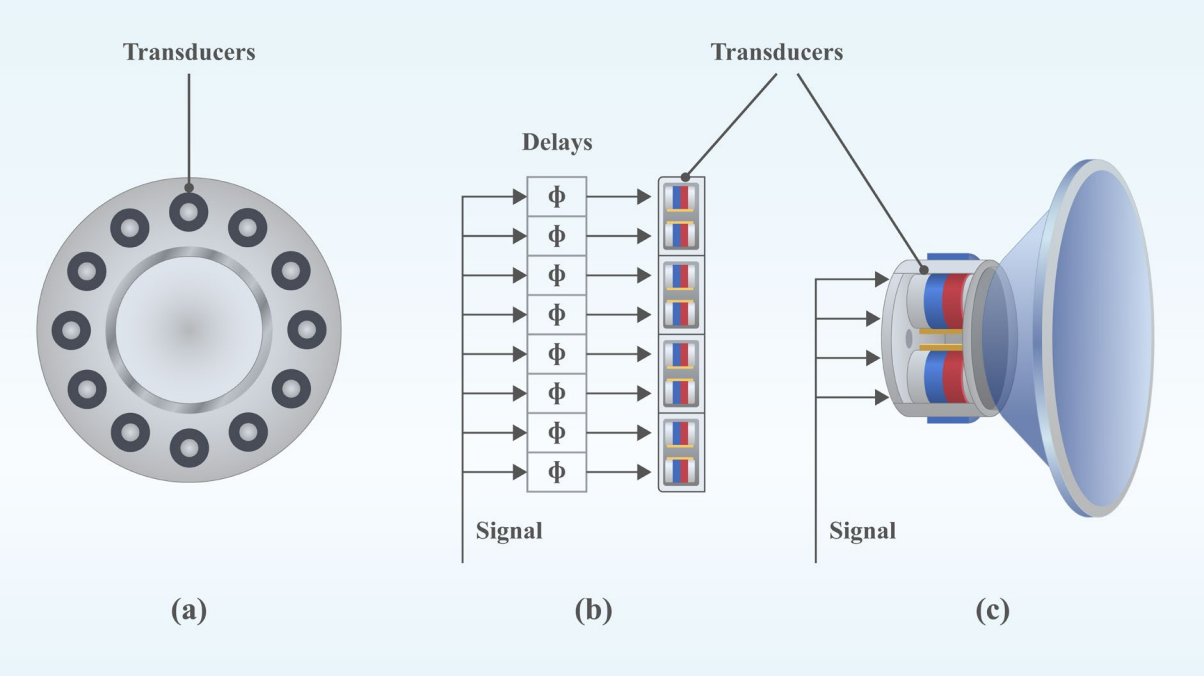


Fig. 2-12. Various parametric speaker array configurations: (a) Ring-shaped transducer arrangement, (b) Phase-deviation array, (c) Horn-augmented structure [1], [66].

A key factor in understanding parametric array loudspeaker directivity is the array aperture, which is equal to the array length. The frequency bandwidth $f_{\min}$ to $f_{\max}$ for maximally directional audio output is empirically constrained by the speed of sound in air ($c \approx 342$ m/s at sea level) [67]:

$$\frac{c}{\left(\frac{1}{2}\right)\text{aperture}} \leq f_{\min} \leq f_{\max} \leq \frac{c}{2\ \text{spacing}} \tag{2-9}$$

Where:

- Aperture size should be at least twice the largest wavelength to maintain collimated sound beams.
- Transducer spacing should not exceed half the wavelength of the highest frequency to prevent undesirable diffraction effects.
- The effective aperture of an ultrasonic transducer array is significantly larger than its actual length, enabling highly focused sound projection even with a compact array design.

Designing array directivity involves several trade-offs. Lower frequencies require larger apertures to maintain beam focus, while higher frequencies demand tighter spacing to avoid grating lobes. Narrower beams improve targeting but reduce the effective bandwidth. These

principles guide the design of efficient PAL arrays for applications such as spatial audio and ANC.

2.4 Acoustic Field Characteristics and Performance Metrics

2.4.1 Near-Field and Far-Field Characteristics

To accurately model the spatial distribution of sound from parametric array loudspeakers, it is crucial to distinguish between near-field and far-field acoustic characteristics (Fig. 2-13). This distinction not only affects the interpretation of beam-forming behaviour but also influences the selection of appropriate modelling techniques for predicting sound pressure and directivity patterns [68], [69].

The Fresnel distance (also known as the near-field to far-field transition distance or Rayleigh distance) is the characteristic range from an acoustic source beyond which the sound field transitions from the near-field (Fresnel zone) to the far-field (Fraunhofer zone). Beyond the Fresnel distance, the wavefront fully develops and becomes less dependent on transducer-specific characteristics. In this region, sound pressure decays according to the inverse-square law. This phenomenon can be described in terms of the acoustic wavelength ‘ λ’ and piston radius ‘a’, with general distances denoted as ’r’ and along the transducer axis as ‘z’. Beyer and Letcher acknowledged the presence of Fresnel and Fraunhofer diffraction, introducing a conceptual distinction between the near-field and far-field at $z=a^2/\lambda$ [70].

Building upon this, Blackstock extended the definition of the far-field to $z>\pi a^2/\lambda$ and $z\gg a$, recognizing the significance of the quantity $\pi a^2/\lambda$ as a marker that delineates the transition from the near-field to the far-field [71]. However, these criteria are not exhaustive. Alternative approaches, such as modelling the near-to-far-field transition via the Rayleigh integral, have also been proposed [72]. Another study further divides the far-field into the Westervelt far-field and the inverse-square law far-field [73]. The latter is characterized by sound pressure being inversely proportional to the distance from the source, adhering to the principles of the inverse square law. It represents the farthest region from the sound source where the sound is perceptible. In this work, we focus on modelling far-field directivity for PALs, particularly within the inverse-square law regime.

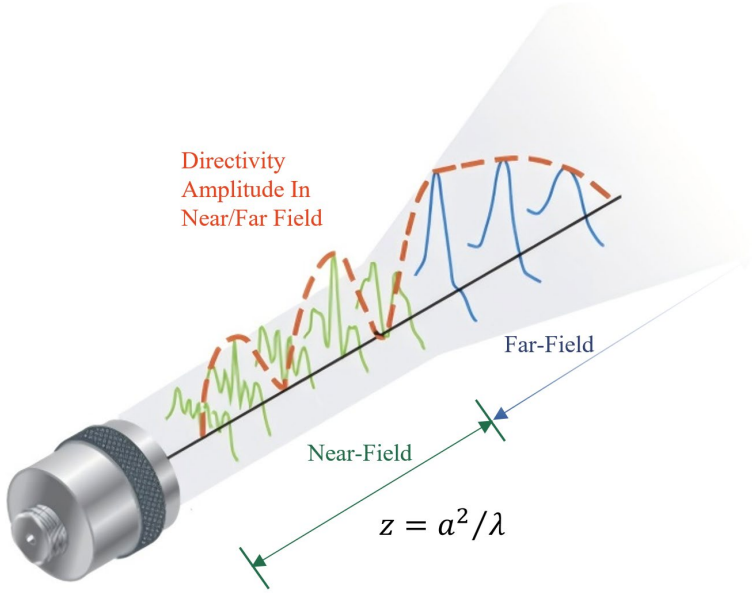


Fig. 2-13. Conceptual illustration of near-field amplitude modulation and far-field beam profile in a parametric acoustic array [72].

2.4.2 Attenuation Coefficient

To accurately model sound propagation in air, especially at ultrasonic frequencies, it is essential to account for the attenuation coefficient, which quantifies the decay of acoustic amplitude with distance. This coefficient is governed by physical mechanisms such as viscous losses, thermal conduction, and molecular relaxation, and is highly frequency dependent. Accounting for attenuation is critical for predicting the sound field generated by PALs. As sound travels through air, its amplitude decreases due to thermo-viscous losses and molecular relaxation, both of which convert acoustic energy into heat. These effects are particularly significant at ultrasonic frequencies. This study adopts a closed-form attenuation model incorporating both classical and relaxation-based absorption, allowing frequency-dependent damping to be explicitly represented [74], [75].

$$\alpha = f^2 \left[1.84\times10^{-11}\frac{p_r}{p_a}\left(\frac{T}{T_0}\right)^{\frac{1}{2}} + \left(\frac{T}{T_0}\right)^{-\frac{5}{2}}\left(\frac{0.01275e^{-\frac{2239.1}{T}}}{f_{r,o}+\frac{f^2}{f_{r,o}}} + \frac{0.1068e^{-\frac{3352.0}{T}}}{f_{r,N}+\frac{f^2}{f_{r,N}}}\right)\right] (NP/m) \quad (2\text{-}10)$$

where α is the acoustic absorption coefficient, f is frequency (Hz), p_a is the ambient air pressure in the propagation environment (Pa), p_r=101.325kPa is the reference atmospheric pressure, T is the absolute air temperature (K), and T_0=293.15 K(20˚C) is the standard reference temperature. The terms $f_{r,O}$ and $f_{r,N}$ denote the relaxation frequencies of oxygen and nitrogen (Hz), respectively. Units: α in Np/m (neper per meter), f and f_r, in Hz , pressures in Pa , and temperatures in K [76].

$$f_{r,O}=\frac{p_a}{p_r}\left(24+4.04\times10^4\mathrm{h}\frac{0.02+\mathrm{h}}{0.391+\mathrm{h}}\right),$$
$$f_{r,N}=\frac{p_a}{p_r}\left(\frac{T}{T_0}\right)^{-\frac{1}{2}}\left(9+280\mathrm{h}_r^{-4.17\left[\left(\frac{T}{T_0}^{-\frac{1}{3}}-1\right)\right]}\right), \tag{2-11}$$

In the Equation (2-11), h is the absolute humidity, which can be derived from the relative humidity h_r and the saturated water vapour pressure p_{sat} by the following equation:

$$h=h_r\frac{p_{\mathrm{sat}}}{p_r}\frac{p_a}{p_r} \tag{2-12}$$

p_{sat} is given by the modified Tetens Equation as:

$$p_{\mathrm{sat}}=0.6112\times\exp\left(\frac{17.67T}{T+243.5}\right) \tag{2-13}$$

Fig. 2-14 shows the variation of sound attenuation coefficients with frequency from 0 to 100kHz at different relative humidities (RH) at 283.15 K(10˚C) and 313.15 K (40˚C), atmospheric pressure (p_r). As depicted, the attenuation coefficients increase with frequency, indicating greater sound energy loss at higher frequencies. In dry air (RH = 0%), water-vapour–assisted vibrational relaxation of O_2 and N_2 is essentially absent, so the relaxation term in atmospheric absorption is negligible and the loss reduces to classical viscous/thermal mechanisms; consequently the attenuation coefficient α is minimal, whereas increasing humidity enhances relaxation absorption in the ultrasonic band [74], [76]. A small non-monotonicity around RH ≈20% at frequency 3 kHz appears where the relaxation frequency passes through the band; for fixed $f_r f_a{}^2/\left(f_r{}^2+f_a{}^2\right)$ peaks at $f_a\approx f_r$, yielding a local maximum and minor curve crossings.

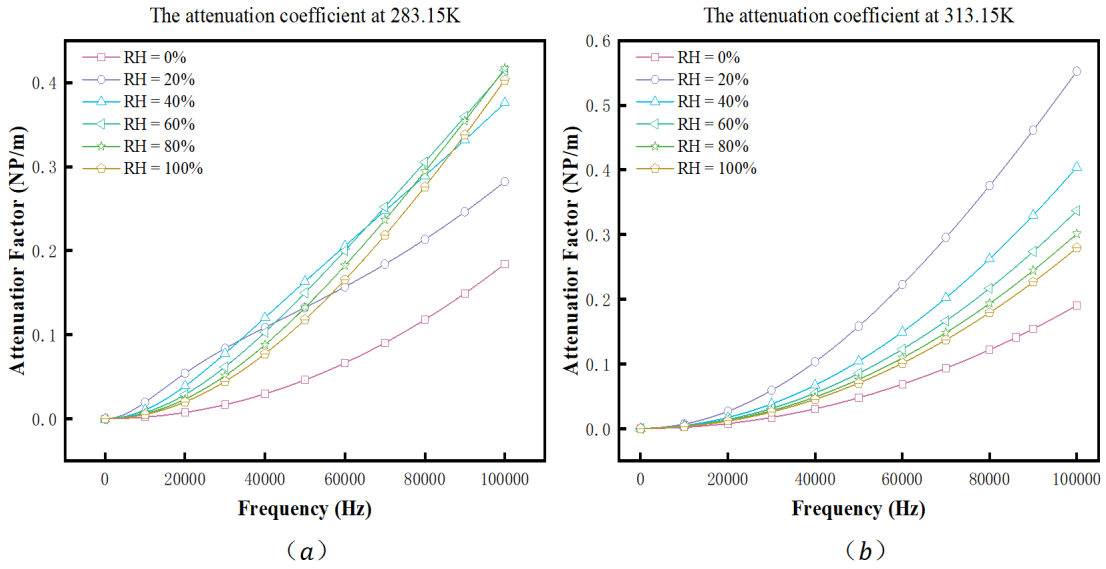


Fig. 2-14. The Frequency response of the attenuation coefficient α at 283.15 K (a) and 313.15 K (b), illustrating the variation in sound attenuation across a frequency range from 0 to 100 kHz.

2.4.3 Directivity of Parametric Speaker

Directivity describes the angular dependence of a loudspeaker's radiated sound field and is a key metric of spatial performance. For PALs, directivity is particularly pronounced because ultrasonic carriers and nonlinear self-demodulation generate sharply focused audible beams with minimal spillover, unlike conventional loudspeakers that typically exhibit broader radiation patterns [101], [102]. In practice, PAL directivity is measured by rotating a microphone at a fixed radius around the source and recording the sound pressure level (or SPL) over angle; the results are plotted in polar or Cartesian form to reveal main-lobe width, side lobes, and off-axis behaviour [103].

In experiments, the SPL-based directivity is defined as a relative level in decibels,

$$D_{\mathrm{SPL}}(\varphi)=L_p(\varphi)-L_p(\varphi_{\mathrm{ref}}) \tag{2-14}$$

where $L_p(\varphi)$ is the measured sound pressure level (in dB) at angle φ, and φ_{ref} is typically the direction of maximum response (e.g., on-axis). Because L_p is already in decibels, the subtraction directly yields a relative directivity in dB.

For theoretical modelling (e.g., FEM), the pressure-amplitude directivity is often defined via a normalized complex pressure (dimensionless),

$$D_p(\varphi)=20\log_{10}\left(\frac{|p(\varphi)|}{|p(\varphi_{\text{ref}})|}\right) \tag{2-15}$$

where $p(\varphi)$ is the (audible) acoustic pressure amplitude at angle φ. p_{rms} is the measured (root-mean-square) sound pressure and the reference in air is $p_{\text{ref}}=20\mu$ Pa [71]. Equations (2-14) and (2-15) are consistent: both express the directional response relative to a reference direction, with (2-15) providing the theoretical counterpart to the experimental SPL definition. In PALs, the observed audible-band pattern reflects the combined effect of the primary ultrasonic beam and nonlinear propagation, which together shape the far-field directivity and its side lobe structure.

2.5 Modelling Strategies for PAL Far-Field Directivity

Accurately modelling the far-field directivity of PALs needs a balance of accuracy, speed, and scalability. The reason is the nonlinear nature of sound propagation and the complex system geometry. Current methods fall into three groups. Analytical models come from nonlinear wave equations. Numerical models use finite element modelling (FEM). Data-driven models use deep learning. This section compares these strategies and evaluates their strengths, limits, and suitability for practical PAL design.

2.5.1 Westervelt-Based Directivity Analytical and Numerical Models

Accurate prediction of PAL sound-field directivity relies on nonlinear acoustic models grounded in basic physical laws. As illustrated in Fig. 2-15, the modelling starts from conservation of mass, conservation of momentum, and the equation of state. Together these describe sound in a fluid. From this basis, researchers derive several nonlinear wave equations. Each equation strikes a different balance between physical fidelity and computational cost..

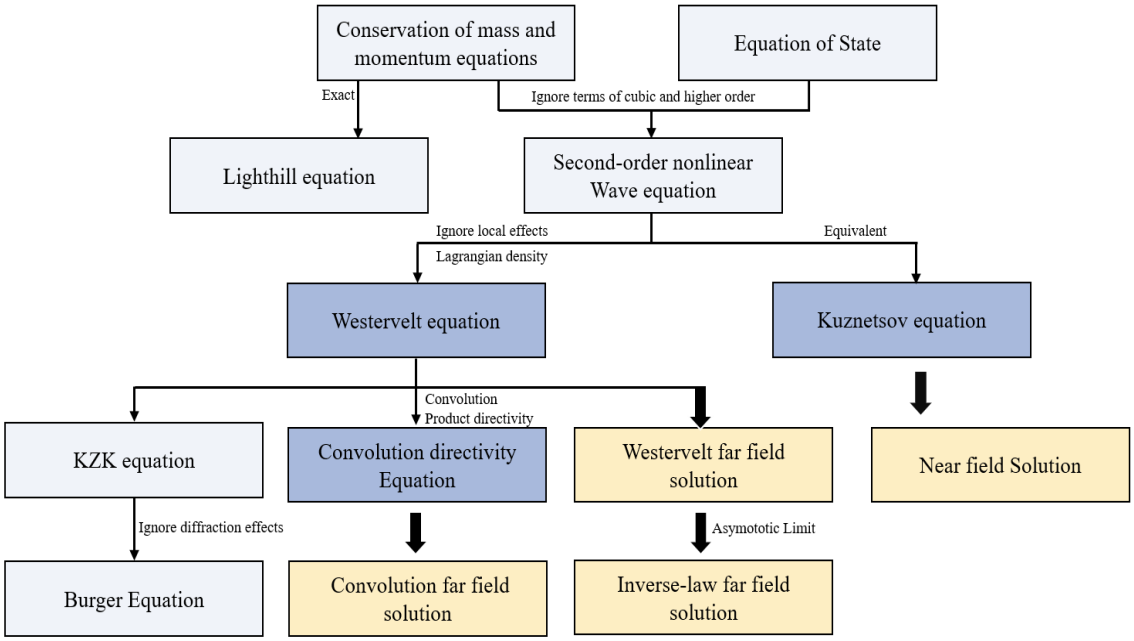


Fig. 2-15. Nonlinear acoustic equation framework for far-field modelling of PALs.

The earliest comprehensive model is the Lighthill equation. It comes from fluid dynamics and models sound generation in turbulent flows. It provides the base for later work in nonlinear acoustics [77].

Then in 1963, the Westervelt equation which is the most widely used nonlinear wave model for PALs invited. It describes finite-amplitude sound in weakly nonlinear media. It adds second-order nonlinearity and diffusive loss. It is well suited to modelling the secondary field produced by self-demodulation in air [4].

From the Westervelt equation, researchers derive simpler forms from Westervelt. The Khokhlov–Zabolotskaya–Kuznetsov (KZK) equation uses a paraxial approximation (KZK) applies a paraxial approximation [78], [79]. It assuming that the primary beam is narrow and forward-propagating. It incorporates nonlinearity, diffraction, and absorption, and is effective in modelling narrow and forward-propagating sound beams [80].

The Burgers [81] equation represents a further simplification of the KZK [82] model, neglecting diffraction entirely. It is typically used in one-dimensional analysis to capture the nonlinear steepening and shock formation in wave propagation [83].

For efficient far-field directivity modelling, the convolution model offers a low-complexity alternative. Derived from the Westervelt equation under far-field assumptions, it represents the secondary field as a convolution between the primary ultrasonic beam and a nonlinear kernel. While it predicts the main lobe accurately, it tends to underrepresent side lobes and array-induced artifacts [84].

The Kuznetsov equation, although mathematically like Westervelt's formulation, uses a slightly different representation and is commonly used in theoretical derivations of nonlinear wave motion [85]. It provides a complete description of nonlinear propagation but is rarely used directly in PAL simulations due to its computational complexity.

Together, these models form a practical ladder for PAL directivity. Westervelt sits at the top for full-physics accuracy; KZK and Burgers simplify the physics when speed is needed; and convolution-based methods give fast far-field estimates that are convenient for large design sweeps and steering studies. Fig. 2-15 summarises how they trade accuracy for computation across this hierarchy.

2.5.2 Acoustic Far-Field Calculation by FEM and Deep Learning

Accurate far-field prediction of PALs requires computational models that can capture the nonlinear acoustic interactions driving secondary sound field formation. While analytical approaches offer theoretical insight, they often miss complex propagation and energy loss in parametric sound. For this reason, researchers develop numerical models. These models simulate parametric-speaker fields under realistic conditions with higher precision [86].

The Westervelt equation is the most used numerical framework for PALs. It includes second-order nonlinearity and dissipative effects. It predicts directional secondary fields with good accuracy. However, its application involves solving second-order partial differential equations with nonlinear source terms, which demand advanced numerical solvers and significant computational resources [87].

FEM suits high-fidelity simulation. It can include nonlinear terms and use adaptive meshing. It can predict fields even with strong nonlinearity and steep wavefronts. For example, Wong et al. (2007) used FEM to simulate CMUTs and showed it can model complex acoustic coupling [88]. In contrast, methods like the Finite-Difference Method (FDM) are often preferred for simpler, homogeneous media and time-domain simulations due to their lower computational

cost. Okita et al. (2011) utilized FDM for high-intensity focused ultrasound (HIFU) simulations in medical applications, highlighting its efficiency in specific contexts [89]. However, for applications such as parametric speakers, even just modelling boundary effects without nonlinear interactions, FEM remains the preferred approach.

FEM is also common in wider acoustics, It models reflections and absorption in concert halls and studios and supports better room design [90]. Taraldsen (2001) demonstrated the effectiveness of FEM in solving nonlinear medical ultrasound problems, providing insights that also apply to parametric speaker modelling [91]. Despite its higher computational demands, the flexibility and accuracy of FEM make it a superior choice for simulating the sound field of parametric speakers, particularly in predicting far-field directivity. As computational power continues to advance, FEM remains a fundamental tool in acoustic engineering, facilitating more precise and efficient parametric speaker designs.

However it still too slow, traditional methods for optimizing parametric speaker arrays, such as Genetic Algorithm (GA) and PSO, rely on computationally expensive simulations like FEM [92]. While these approaches offer accurate solutions, their high-cost limits real-time applications and large-scale optimisation. To address this, generative models like GANs present a promising alternative by learning complex mappings between array configurations and their resulting directivity patterns, significantly reducing computational requirements [93].

GANs have shown strong performance in domains such as image synthesis and audio signal enhancement, with models like WaveGAN and SEGAN demonstrating realistic waveform generation and noise suppression [94]. Although these models primarily target one-dimensional signals, recent studies have applied GANs to physics-informed tasks such as fluid dynamics, where they significantly reduce simulation cost while preserving spatial accuracy [95]. These findings suggest that GANs can be extended to nonlinear acoustic modelling, offering a data-driven alternative to solve the Westervelt equation efficiently.

To systematically compare modelling strategies for predicting PAL far-field directivity, Table 2-2 summarizes three representative approaches, analytical, numerical, and data-driven, along with their typical strengths and limitations. Analytical methods are fast and interpretable with closed or semi-closed forms, but their simplifying assumptions limit accuracy in complex or off-axis scenarios. Numerical methods deliver high-fidelity predictions, handle complex geometries and boundaries, and can include nonlinearity and attenuation, yet they are

computationally expensive and less scalable. Data-driven models provide fast inference, are scalable and optimizer-friendly, but remain data-dependent and offer weaker physical guarantees. These trade-offs motivate a hybrid workflow in which high-fidelity simulations inform learned models, combining the accuracy of numerical methods with the efficiency and scalability of data-driven prediction for PAL directivity estimation.

Table 2-2. Representative Approaches for PAL Directivity Prediction: Methods, Representative References, and Typical Applications.

Category	References	Limitations	Strengths
Analytical	Westervelt [18], KZK [82], Burgers [81]	Oversimplified, low accuracy	Fast; interpretable
Numerical	FEM [96], FDTD [97], BEM [98],	Computationally expensive, Scalability limited	High-fidelity, complex geometry and boundaries, nonlinear attenuation
Data-driven	WaveGAN[6] SEGAN[94]	Data-dependent, weak physical guarantees	Fast inference, scalable, optimizer-friendly

2.6 Summary

This chapter has reviewed the foundations needed to model and design PALs. We outlined the nonlinear acoustics that govern ultrasonic carriers and difference-frequency generation, then connected these principles to practical system choices: transducer layouts, modulation schemes, and beamforming strategies for directional projection. Because drive electronics shape the usable bandwidth and modulation fidelity, we also noted relevant electrical aspects, including amplifier and transducers. Building on this context, we compared three modelling families analytical (Westervelt/KZK/Burgers), numerical (FEM), and data-driven (GAN) with attention to what each deliver in terms of interpretability, spatial fidelity, runtime, and ease of integration into a design workflow.

The comparison clarifies the trade-offs. Analytical methods are fast and interpretable, but they rely on simplifying assumptions that limit their ability to represent nonlinear loss, edge effects, and grating-lobe behaviour in steered or non-uniform arrays. FEM achieves high spatial fidelity

and accommodates complex boundaries; however, its computational cost hinders routine design sweeps, optimisation, and rapid exploratory studies. GAN-based models provide near-instant inference and scale well to large design spaces, yet their reliability depends on the coverage of the training data and they offer weaker explicit physical guarantees.

From this review, we can find that there is no widely adopted approach that simultaneously delivers FEM-level spatial fidelity for main-lobe and side/grating-lobe prediction under realistic steering and layouts and the design-time efficiency required for parameter sweeps and optimisation. In practice, PAL development needs a modelling route that preserves the physical character of the sound field, scales to large design studies, and is evaluated with consistent, peak-normalized, measurement-referenced metrics on matched angle samples.

These observations motivate a hybrid direction. The next chapter introduces a dual-convolution, physics-aware modelling approach that embeds array geometry and spatial impulse-response information to better capture far-field structure including side lobes and grating lobes while remaining lightweight enough for broad design sweeps and optimisation. Subsequent chapters evaluate this approach alongside FEM and measurement under a consistent protocol and show how it can support steerable PAL configuration with the accuracy, efficiency, and reproducibility that real-world workflows require.

Chapter 3

Grating Lobe Prediction Using a Dual Convolution Model for PALs

3.1 Introduction

Traditional approaches to modelling the far-field characteristics of PALs, such as analytical solutions based on the Westervelt equation and numerical simulations using FEMs, often face significant computational challenges. While conventional Westervelt-based convolution models are efficient in capturing main lobes, however, they do not resolve the precise grating lobes when the array layout or steering changes. limits their practical use in PAL design. To address this, we propose a Dual Convolution Model to predict the grating lobes in the directivity patterns of PALs.

Here, a novel approach integrates a quasi-linear solution of the Westervelt equation with the two-dimensional (2D) convolutional model for directivity analysis. To enhance directivity prediction accuracy, the derived expression is augmented convolution with the phased-array eigen-directivity, yielding the modified convolutional formulation model which is called the dual convolution model. Convolution can be used to model the superposition effect of the sound waves in space. The enhanced model not only improves main lobe accuracy but also extends predictions to side lobes and grating lobes. Measurements of the directivity of rectangular PALs are compared with the proposed modified convolution model, validating the accuracy in predicting directivity and efficacy to capturing side-lobes and grating lobes. The proposed model can provide valuable support for future investigations into the design and implementation of the far-field of PALs.

In this Chapter First, we present the theoretical foundations of the Dual Convolution Model. Then we build the predictive framework from these foundations. Next, we describe the experimental setup and report results, comparing the proposed model with standard convolution methods and FEM. Finally, we show that the model captures grating-lobe characteristics and improves predictive accuracy.

3.2 Westervelt-Equation-Based Dual Convolution Model

In this section, we model the directivity of PALs. First, we derive the standard convolution model and relate its terms to the array geometry and the propagation path. Then, we introduce a two-dimensional predictive model that estimates grating lobes with high accuracy. Next, we extend the derivation to three dimensions to cover more general layouts and steering. In the last part, we explain why the 2D model is sufficient in most practical cases: the principal-plane behaviour dominates, out-of-plane variation is weak over the apertures and angles considered, and the evaluation metrics are taken on those planes.

3.2.1 2D Analysis of analytical method

Fig. 3-1 is a schematic that explains the PAL architecture and demodulation mechanism; the actual FEM geometry and mesh used in simulations are shown in Fig. 4-2 (COMSOL screenshot with PML and a mesh inset). The initial model considers a PAL placed on an infinitely reflective plane, with its three-dimensional (3D) configuration depicted in Fig. 3-1(b). In this setup, the PAL emits ultrasonic beams at distinct frequencies, while the DF component which representing the audible signal is generated via second-order nonlinear interactions in air.

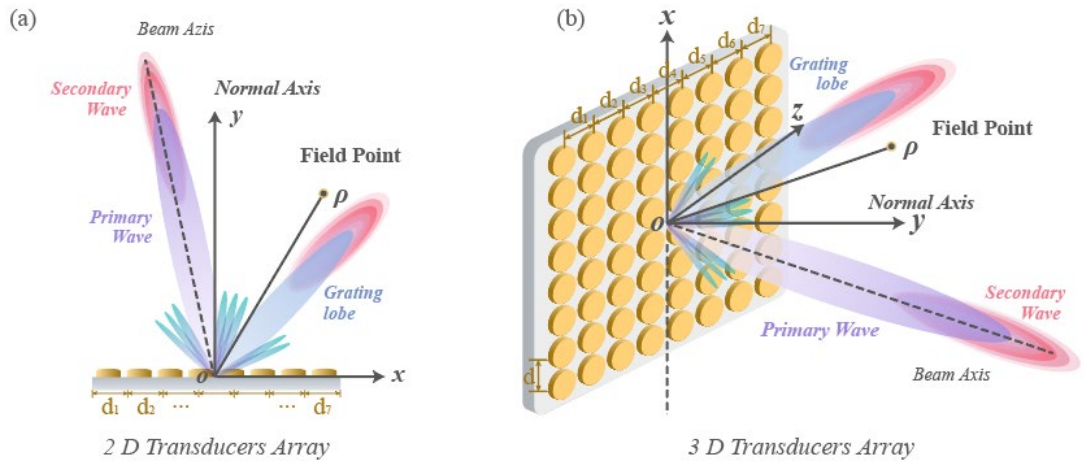


Fig. 3-1. The linear acoustic phased array and corresponding sound field, illustrating (a) an 8-element configuration in the 2D model and (b) an 8×8-element configuration in the 3D model, both depicted with a specified.

Although full 3D modelling can reveal rich acoustic behaviour and enable applications such as medical imaging and complex environmental simulations [99], [100], such scenarios extend beyond the scope of this work. In contrast, the primary focus here is on planar far-field directivity in controlled acoustic environments. For this purpose, a 2D simplification is both computationally efficient and sufficiently accurate. By constraining the problem space, the 2D model in Fig. 3-1(a) retains essential physical characteristics such as beam steering, grating lobe formation, and spatial aliasing while significantly reducing the computational burden.

In a 2D analysis, a Cartesian coordinate system (x, y) and a polar coordinate system (ρ, φ) are centred at the PAL's origin, O. Here, the x-axis' positive semiaxis aligns with the sound radiation direction, with ρ and φ representing radial and azimuthal coordinates, respectively. [81].

In the following, the Westervelt equation will be used to derive the far-field solution of phase loudspeaker array. In the derivation, the convolution model proposed in recent studies [101] is employed to accurately characterize the far-field behaviour. The radiation of the loudspeaker arrays in the far-field is dominated by the Westervelt equation as [102]:

$$\nabla^2 p - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} = -\frac{\delta}{c_0^4}\frac{\partial^3 p}{\partial t^3} - \frac{\beta}{\rho_0 c_0^4}\frac{\partial^2 p^2}{\partial t^2} \tag{3-1}$$

Where ρ_0 represents the ambience air density, $c_0$is the linear speed of sound in air, p denotes the sound pressure, β is the nonlinear coefficient in air, and δ is the sound diffusivity which is linked to the sound attenuation coefficient α which defined in Equation (3-1). The Westervelt equation serves to precisely simulate the PAL under second-order nonlinearities within the far-field. Employing a stepwise approach, it simulates the quasi-linear solution of audio sound, representing as the emission from an infinite virtual source with a defined sound pressure [103]:

$$p_a(\rho) = -\rho_0 w_a \int_0^{2\pi}\int_0^{\infty} q_a(\rho_v) H_0\left(k_a|\rho-\rho_v|\right)\rho_v \, \mathrm{d}\rho_v \, \mathrm{d}\varphi_v \tag{3-2}$$

Here, the virtual source point $\rho_v = (x_v, y_v)$, and $|\rho-\rho_v| = \sqrt{(x-x_v)^2+(y-y_v)^2}$ is the distance between the point ρ and the source point ρ_v. $\omega_a = 2\pi f_a$ is the angular frequency and $f_a = f_2 - f_1$ is the

frequency of the audio. H_0 is the first order Hankel function, $k_a = \omega_a/c_0$ is the wavenumber and $q_a$as the density of the source [103], which can be expressed as:

$$q_a(\rho_v)=\frac{\beta\omega_a}{\mathrm{i}\rho_0{}^2c_0{}^4}p_1{}^*(\rho_v)p_2(\rho_v) \tag{3-3}$$

The acoustic pressure of ultrasound at the virtual source point can be formulated in a manner akin to the Rayleigh integral [103]:

$$p_{\mathrm{j}}(\rho_v)=\frac{p_0}{2}\int_{-a_p}^{a_p}H_0(k_j|\rho_v-\rho_s|)u_j(\rho_s)\mathrm{d}y_s \tag{3-4}$$

Here $a_p$denotes the piston half-width (aperture parameter). The symbol $a$elsewhere refers to the attenuation combination $a=\alpha_1+\alpha_2$. In the far-field case with $\rho\to\infty$, the 0th order Hankel function has an approximation of [26]:

$$H_0(k_a|\rho-\rho_v|)\approx\sqrt{\frac{2}{\mathrm{i}\pi\rho k_a}}\mathrm{e}^{\mathrm{i}k_a[\rho-\rho_v\cos(\varphi-\varphi_v)]} \tag{3-5}$$

where:

$$|\rho-\rho_v|\approx\rho-\frac{\rho\cdot\rho_v}{\rho}=\rho-\rho_v\cos\varphi-\varphi_v \tag{3-6}$$

Considering the scenario of far-field with $\rho_v\to\infty$, Equation (3-4) is expressed as [27]:

$$p_j(\rho_v)=\sqrt{\frac{2}{\mathrm{i}\pi k_j\rho_v}}p_0k_ia_p\mathrm{e}^{\mathrm{i}k_i\rho_v}D_j(\varphi_v) \tag{3-7}$$

Substituting Equation 3-7 into Equation (3-4) yields the source density of the audio sound [101]:

$$q_a(\rho_v)=\frac{2\beta k_ap_0{}^2\sqrt{k_1k_2}a_p{}^2}{\mathrm{i}\pi\rho_0{}^2c_0{}^3\rho_v}\mathrm{e}^{\mathrm{i}(k_2-k_1{}^*)\rho_v}D_1{}^*(\varphi_v)D_2(\varphi_v) \tag{3-8}$$

Applying Equation3-8 into Equation (3-2) we can derive the sound pressure as [103]:

$$p_a(\rho)=\frac{2\beta k_a{}^2 p_0{}^2\sqrt{k_1k_2}a^2}{\pi i\rho_0 c_0{}^2}$$
$$\int_0^{2\pi}\int_0^{+\infty} D_1{}^*(\varphi_v)D_2(\varphi_v)e^{i(k_2-k_1{}^*)\rho_v}H_0(k_a|\rho-\rho_v|)\rho_v\,d\rho_v\,d\varphi_v \tag{3-9}$$

With implementing Equation (3-5) into Equation (3-9) the sound pressure can be expressed as [103]:

$$p_a(\rho)=\frac{4\beta k_a{}^2 p_0{}^2 a_p{}^2}{i\pi}\sqrt{\frac{k_1k_2}{2\pi i k_a\rho}}e^{ik_a\rho}$$
$$\int_0^{2\pi}\int_0^{+\infty} D_1{}^*(\varphi_v)D_2(\varphi_v)e^{i[k_2-k_1{}^*-k_a\cos(\varphi-\varphi_v)]\rho_v}\,d\rho_v\,d\varphi_v. \tag{3-10}$$

We define an integral called Westervelt's directivity as [101]:

$$D_w(\varphi)\equiv(\alpha_1+\alpha_2)\int_0^{+\infty} e^{i[k_2-k_1{}^*-k_a\cos(\varphi-\varphi_v)]\rho_v}\,d\rho_v$$
$$=\frac{1}{1-ik_a\frac{\alpha_1+\alpha_2}{2}{}^{-1}\sin^2\left(\frac{\varphi-\varphi_v}{2}\right)} \tag{3-11}$$

The total attenuation coefficient $\alpha_1+\alpha_2$ is such that $D_w(\varphi)$ has a value of 1 at $\varphi=0$. In this case, Equation (3-10) can be simplified to:

$$p_a(\rho)=\frac{4\beta k_a{}^2 p_0{}^2 a_p{}^2}{i\pi}\sqrt{\frac{k_1k_2}{2\pi i k_a\rho}}e^{ik_a\rho}D_a(\varphi) \tag{3-12}$$

where audio sound directional equation is expressed as:

$$D_a(\varphi)=(D_1{}^*D_2\times D_w)(\varphi)=\int_0^{2\pi} D_1^*(\varphi_v)D_2(\varphi_v)D_w(\varphi-\varphi_v)d\varphi_v \tag{3-13}$$

In the 2D piston model, the velocity distribution of the ultrasound is uniform $u_j(\rho_s)=1, \rho_s\le a_p$, where the directivity of the corresponding ultrasound is $D_j(\varphi)=\mathrm{sinc}(k_j r\cos\varphi)$. Here, sin$c$ is the first order spherical Bessel functions of class 0.

The symbol ×signifies the linear convolution operator. Equation (3-13), commonly known as the convolution model, demonstrates that the audio sound's directivity results from the convolution of the Westervelt directivity of the ultrasound and the product directivity. The model does not account for nonlinear interactions in the near-field and neglects transducer spacing effects, leading to inaccuracies in predicting audio directivity.

The parametric speaker, regarded as a phased array, can be analysed using the phased array model to a reasonable degree in predicting its far-field characteristics. Moreover, its main lobes can be regulated via the phase difference between transducers. The results of a uniform linear array in Equation 3-14 for the far-field polar response, at a frequency which wavenumber is k. From 3-14 onward, we use θ for the principal-plane steering angle; φ and φ_v remain azimuthal angles.

$$P(\theta_k)=A_0 e^{-j(0)\,kd\sin\theta}+A_1 e^{-j(1)\,kd\sin\theta}+A_2 e^{-j(2)\,kd\sin\theta}+\cdots+A_{N-1} e^{-j(N-1)\,kd\sin\theta} \tag{3-14}$$

Where d is the distance between each transducer, and the wavenumber k is given by:

$$k=\frac{2\pi}{\lambda}=\frac{\omega}{c_0}=\frac{2\pi f}{c_0} \tag{3-15}$$

This can be rewritten as:

$$P(\theta)=\sum_{n=0}^{n} A_n e^{-jnkd\sin\theta} \tag{3-16}$$

To implement the deflection, an additional weight matrix is required:

$$W_p(\theta_v)=\sum_{n=0}^{n} e^{-jnkt\tau} \tag{3-17}$$

Where $\tau = d\sin\theta/c$ is the delay between each transducer, so that, the direction of the phase array can be set as:

$$D_{\mathrm{p}}(\theta,\theta_v)=W_p(\theta)\cdot P(\theta_v)=\sum_{n=0}^{n} e^{-j2nkd\sin\theta} e^{-jnkt\tau} \tag{3-18}$$

By convolving the phased array's directivity with the convolution model obtained from the above equation, a directivity model with the nature of the phased array can be obtained, which is called the Dual convolution model in this thesis:

$$\widetilde{D_P}(\varphi)=\left(D_1 \,{}^{*}D_2\times D_w\right)\times D_{\mathrm{p}}(\varphi) \tag{3-19}$$

To determine the overall computational complexity of the algorithm described by the equations, we need to consider the individual complexities of each component of the computation and how these scale with respect to the input parameters.

The complexity of this algorithm shows that the Equation (3-20) consists of $D_1 \,{}^{*}D_2\times D_w$, $D_{\mathrm{p}}(\varphi)$. The Equation $D_{\mathrm{p}}\left(\varphi,\varphi_v\right)$ involves a summation from n=0 to N, each involving complex exponential calculations dependent on θ and τ. The complexity for this summation is $O(N)$. Assuming $D_1 \,{}^{*}D_2\times D_w$and $D_{\mathrm{p}}(\varphi)$ are evaluated at M discrete values of φ, and convolution is implemented using Fast Fourier Transform (FFT). The overall computational complexity of the algorithm can thus be summarized as:

$$C= O(M\ log\ M) \tag{3-20}$$

This expression succinctly encapsulates the scaling behaviour with respect to the number of summation terms N and the convolution complexity involving M discrete angles, providing a clear picture of the algorithm's computational demands.

3.2.2 3D Analysis of analytical method

In 2D we adopt Cartesian (x, y) and polar (ρ, φ) coordinates centred at O, with the positive x-semiaxis aligned with the radiation direction. In 3D we use Cartesian (x, y, z) and spherical (r, θ, φ) centred at O, with the z-axis along the radiation direction; r is radial distance, θ the polar (colatitude) angle from +z, and φ the azimuth.

For tractability, we first present the 2D derivation; the 3D counterpart is outlined below. At the audio angular frequency $\omega_a=\omega_2-\omega_1$, incorporating absorption into a wavenumber $k_a=\omega_a/c_0$, the Westervelt equation reduces to an in homogeneous Helmholtz equation with a quadratic source [103]:

$$(\nabla^2+k_a^2)p_a(\mathbf{r})=S_a(\mathbf{r}),\ S_a(\mathbf{r})=C_a p_1^*(\mathbf{r})p_2(\mathbf{r}), \tag{3-21}$$

Where C_a is Quadratic-source constant $C_a \propto \beta\omega_a^2/(\rho_0 c_0^4)$ with β the nonlinear coefficient. Using the free-space *Green's* function $G_a(\mathbf{r})=e^{ik_a|\mathbf{r}|}/(4\pi|\mathbf{r}|)$, the DF field approximation is:

$$p_a(\mathbf{r})=\iiint_{R^3} G_a(\mathbf{r}-\mathbf{r}_v)S_a(\mathbf{r}_v)d^3\mathbf{r}_v \tag{3-22}$$

Far-field (Fraunhofer) approximation. For $r=|\mathbf{r}|\to\infty$ and observation direction $\hat{\mathbf{r}}=(\theta,\varphi)$ [101],

$$|\mathbf{r}-\mathbf{r}_v|\simeq r-\hat{\mathbf{r}}\cdot\mathbf{r}_v,\ G_a(\mathbf{r}-\mathbf{r}_v)\simeq\frac{e^{ik_a r}}{4\pi r}e^{-ik_a\hat{\mathbf{r}}\cdot\mathbf{r}_v}. \tag{3-23}$$

Hence

$$p_a(\mathbf{r})\simeq\frac{e^{ik_a r}}{4\pi r}\iiint S_a(\mathbf{r}_v)e^{-ik_a\hat{\mathbf{r}}\cdot\mathbf{r}_v}d^3\mathbf{r}_v \tag{3-24}$$

Here $\hat{r}$ denotes the observation direction with spherical angles (θ,φ). Primary (ultrasonic) far-field at the virtual source. For $j\in\{1,2\}$,

$$p_j(\mathbf{r}_v)\simeq A_j D_j^{(3D)}(\Omega_v)\frac{e^{ik_j r_v}}{r_v},\ \Omega_v=(\theta_v,\varphi_v) \tag{3-25}$$

where $D_j^{(3D)}(\theta,\varphi)$ denotes the 3D primary directivity (element pattern times array factor if applicable), and A_j is a constant for fixed drive.; $\Omega_v=(\theta_v,\varphi_v)$is the virtual-source direction; $\alpha_a$is the DF attenuation coefficient. Therefore,

$$S_a(\mathbf{r}_v)=C_a p_1^*(\mathbf{r}_v)p_2(\mathbf{r}_v)\propto D_1^{(3D)*}(\Omega_v)D_2^{(3D)}(\Omega_v)\frac{e^{i\Delta k r_v}}{r_v^2},\ \Delta k=k_2-k_1 \tag{3-26}$$

We define γ is the angle of observation direction $\hat{\mathbf{r}}=(\theta,\varphi)$ and virtual-source direction $\Omega_v=(\theta_v,\varphi_v)$. With $d^3\mathbf{r}_v=r_v^2 dr_v\, d\Omega_v$ a so that $\hat{\mathbf{r}}\cdot\mathbf{r}_v=r_v\cos(\gamma)$, the attenuation of DF is α_a. substituting (3-26) into (3-24) gives

$$p_a(\mathbf{r})=\frac{e^{ik_a r}}{r}\int_{S^2} D_1^{(3D)*}(\Omega_v)D_2^{(3D)}(\Omega_v)\underbrace{\int_0^{\infty} e^{-\alpha_a r_v}e^{i[(\Delta k)-k_a\cos(\gamma)]r_v}dr_v}_{D_w^{(3D)}(\gamma)}d\Omega_v \tag{3-27}$$

This defines the 3D Westervelt directivity function

$$D_w^{(3D)}(\gamma)=\int_0^{\infty} e^{-[\alpha_a-i(\Delta k-k_a\cos\gamma)]r_v}dr_v=\frac{1}{\alpha_a-i(\Delta k-k_a\cos\gamma))} \tag{3-28}$$

After normalization $\bar{D}_w^{(3D)}(\varphi)\equiv D_w^{(3D)}(\varphi)/D_w^{(3D)}(0)$ (so $\bar{D}_w^{(3D)}(0)=1$), the audio directivity is a spherical convolution (*) denotes convolution over the unit sphere):

$$D_a^{(3D)}(\theta,\varphi)=\left[D_1^{(3D)*}D_2^{(3D)}\right]\times\bar{D}_w^{(3D)}(\theta,\varphi). \tag{3-29}$$

Here "×" denotes convolution over the unit sphere, and "*" denotes complex conjugation. When phased-array steering is present with a 3D array factor $D_p^{(3D)}(\theta,\varphi)$, the total (dual-convolution) directivity is

$$\tilde{D}_P^{(3D)}(\theta,\varphi)\propto\left[D_1^{(3D)*}D_2^{(3D)}\right]\times\bar{D}_w^{(3D)}\times D_p^{(3D)}(\theta,\varphi). \tag{3-30}$$

Equations (3-30) - (3-31) are the exact 3D counterparts of 2D convolution (3-19) and dual-convolution (3-26); the line convolution in φ is replaced by a spherical convolution in (θ,φ).

For a uniform linear array (elements along x) with weight/delay w_n, the 3D array factor is

$$D_p^{(3D)}(\theta,\varphi)=\sum_{n=0}^{N-1} w_n e^{-jknd\sin\theta\cos(\varphi-\varphi_0)}. \tag{3-31}$$

On the principal plane $\theta=\pi/2$,

$$D_p(\varphi)=\sum_{n=0}^{N-1} w_n e^{-jknd\cos(\varphi-\varphi_0)}, \tag{3-32}$$

which matches the 2D array factor used in (3-17)-(3-19). Substituting (3-31) and using (3-28) in (3-32) yields the dual-convolution on φ :

$$\widetilde{D}_P^{(3D)}\left(\frac{\pi}{2},\varphi\right)=\left[\left(D_1^{*}D_2\right)\times\bar{D}_w^{(2D)}\times D_p\right](\varphi), \tag{3-33}$$

As the same functional form as (3-25) in 2D. The only difference between 3D and 2D fields is the geometric spreading factor $e^{ik_a r}/r$ versus $e^{ik_a\rho}/\sqrt{\rho}$, which vanishes under normalized plots.

Let M be the number of azimuth samples in 2D and L the spherical-harmonic bandlimit in 3D.

- 2D (cf. (3-20)): FFT convolution over M angles costs $O(M \log M)$.
- 3D: spherical convolution via spherical-harmonic transforms typically costs $O(L^3)$ (or $O(L^2 \log L)$ with specialized algorithms) with a much larger constant.

A full 3D treatment is only necessary when

- Significant polar-angle variation is required (elevation scanning or tall apertures);
- Planar arrays/3D arrays with non-negligible vertical aperture;
- Near-field regimes (Fraunhofer invalid);
- Strongly asymmetric boundaries cause $D_j^{(3D)}$ to vary sharply with θ. None applies to our linear-array, far-field, principal-plane evaluation; thus, 2D modelling yields the correct normalized principal-plane directivity at a fraction of the cost.

3.3 Experimental Setup for Verifying the Dual-Convolution Model

3.3.1 PAL Configuration and Signal Generation

In light of the theoretical analysis in chapter 3.2, the normalized directivity of a parametric array loudspeaker in the far-field can be accurately represented by the 2D convolution model. The 3D formulation reduces to the 2D expression when restricted to the principal plane, and the additional computational burden of full 3D modelling is unnecessary for evaluating azimuthal beam patterns. Based on this equivalence, the first experimental validation was designed to measure the far-field directivity of a parametric array loudspeaker along a 2D sampling plane. This approach allows a direct comparison with the 2D convolution model, providing a reliable verification of its predictive capability.

The experimental system uses a uniform 8 × 8 ultrasonic array (Fig. 3-2). Each element is a piezoelectric transducer mounted on a rigid baffle with consistent spacing. Each channel is

driven with double-sideband amplitude modulation (DSB-AM) to generate the audio difference-frequency in air. An Arduino Mega provides eight digital carrier waveforms with programmable phase and duty settings; The audio baseband is scaled and DC-biased to form a non-negative envelope, and the firmware multiplies this envelope with the carrier to produce a non-suppressed DSB-AM signal (carrier plus sidebands) at the chosen ultrasonic frequency. The modulated bitstreams feed Class-D output stages built around TC4427 MOSFET gate drivers that drive compact MOSFET half-bridges. A multichannel shield boosts each output to approximately 17 Vpp so that the piezoelectric transducers can be driven independently.

Fig. 3-2. Photograph of the physical transducer array configurations: Uniform 8 × 8 array with fixed 10 mm inter-element spacing.

3.3.2 Measurement Setup and Data Acquisition

Measurements were performed in an anechoic environment at a distance of 2 m from the array, ensuring operation within the far-field region. The directivity was sampled over an azimuthal angular range of 180° with a resolution of 5°, yielding 38 discrete measurement points. This 2D acquisition captures the principal-plane radiation pattern of the parametric speaker while avoiding unnecessary complexity from full 3D scanning. Unlike the Chapter 4 experiments that used a motorised turntable, this setup relied on manual alignment. We calibrated the angle marks before each run and checked them at several reference positions to limit cumulative error. We are upgrading the turntable for future measurements in Chapter 4, so that the same protocol can be repeated with automated rotation and improved repeatability.

The experiment setup uses the PRIMO EM258 Model as a microphone, which was employed to acquire acoustic data. This device can detect sound frequencies exceeding 95 kHz. Key specifications of the PRIMO EM258 include a Signal to Noise ratio of 74 dB with a self-noise level of 20 dB, a sensitivity of -32 dB with a tolerance of +/- 3 dB at 1kHz (where 0dB is equivalent to 1V/Pa), and an operating voltage range of 3V, accommodating a broader range of 1.5-10 V.

The signals were bandpass-filtered to isolate the demodulated difference-frequency component at 4 kHz. from this filtered segment, the RMS pressure p_{rms} was computed, and the corresponding SPL was calculated using Equation (2-14). To ensure consistency and comparability between configurations, all SPL values were normalized by subtracting the peak SPL within each angular sweep, resulting in relative directivity profiles expressed in dB. These profiles were used to generate polar plots for evaluating main lobe orientation, beamwidth, and side lobe characteristics. The complete acquisition and analysis workflow is illustrated in Fig. 3-3.

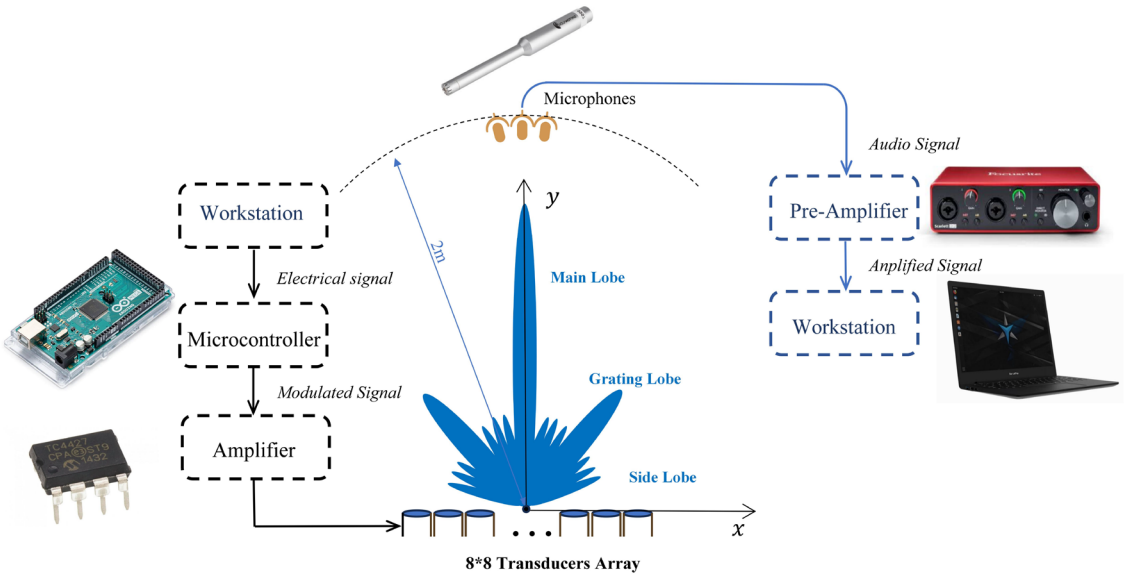


Fig. 3-3. Schematic diagram of the measurement process. The microcontroller generates digital excitation signals, which are amplified and delivered to an 8 × 8 transducer array.

At each angular position, a time-domain waveform was recorded with a microphone and 24-bit interface; only the steady-state window was analysed. Before measurements we performed electrical and acoustic verification of the array:

i. Checked and labelled the polarity (+/-) of every transducer and wiring harness;
ii. Used an oscilloscope to verify, at each driver pin, the carrier frequency, the programmed phase delay, and the peak-to-peak voltage, ensuring channel-to-channel phase alignment and absence of clipping;
iii. After connection, conducted an acoustic sanity check by using a receiver (microphone) to confirm the presence of the ultrasonic carrier and the demodulated audio tone at the measurement radius, and to verify that the main lobe steered to the expected angle under the applied delays.

The primary calibration in this study was the per element polarity, phase, frequency, and drive-level verification described above.

3.4 Results Validation for the Dual-Convolution Model

3.4.1 Dual Convolution Model Verification:

In the conventional convolution model [67], beam steering is implemented by embedding the phased-array phase law directly into the primary directivities D_1 and D_2: the physical aperture is partitioned into short sub-segments, each segment is assigned the appropriate phase, and the contributions are superposed to form a single convolution result. This quasi-continuous treatment is efficient, but it folds the array geometry into the primary fields, which cause:

i. Blurs discrete-element effects;
ii. Makes the result sensitive to how the aperture is segmented;
iii. Can under- or over-estimate grating-lobe levels when element spacing approaches or exceeds $\lambda_c/2$ or when the layout is non-uniform.

The dual-convolution model separates these roles. We retain D_1 and D_2 for the primary element pattern (or measured/analytic single-element response) and apply the propagation kernel as before, then introduce an explicit array factor D_p built from the actual transducer centres and delays. This decoupling lets the model (a) operate on the discrete layout without sub-segmentation, (b) preserve the steering law exactly, and (c) recover grating-lobe positions and levels more faithfully for sparse or non-uniform arrays, while adding only a negligible FFT-level cost. All models were assessed under the same normalisation and angular sampling; results and RMSE statistics are reported in this section.

To validate the proposed dual-convolution model, we performed comparative simulations against a Westervelt-based model and the conventional convolution model at a 40 kHz carrier. For a slender, effectively linear aperture, the principal-plane (2D) solution provides an accurate approximation to the full 3D far-field, so all models were evaluated in 2D and compared with peak-normalised measurements at matched angles (Fig. 3-4).

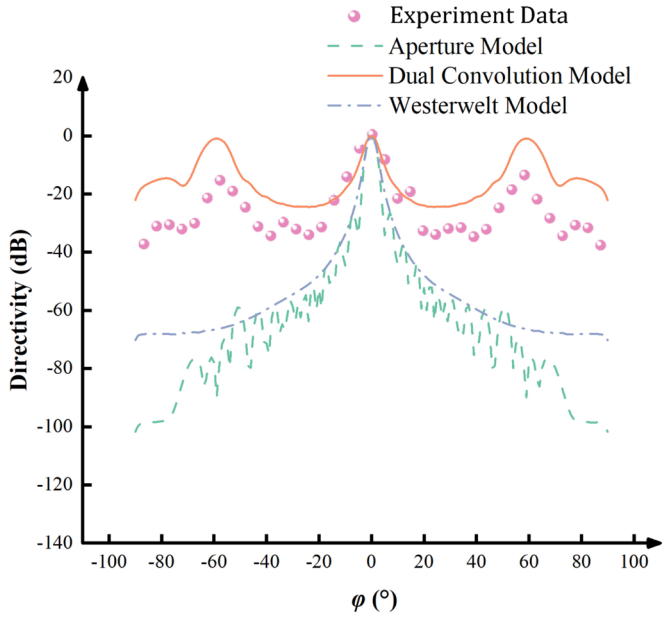


Fig. 3-4. Comparison of experimental data and simulated data for directivity of rectangular deflectable PAL model at 0°deflection.

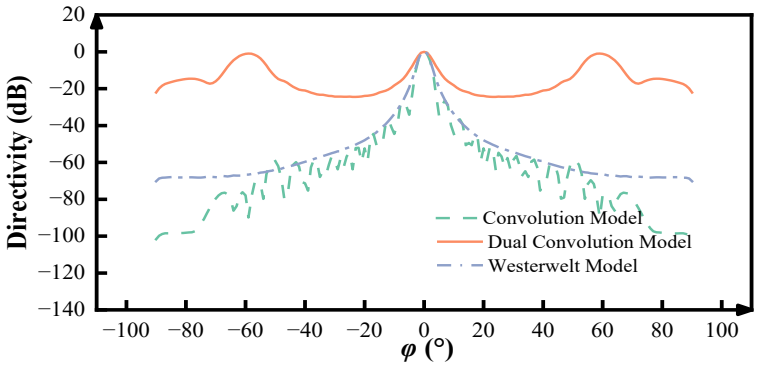
After validating the models, we conducted benchmark simulations to examine parameter effects on grating-lobe formation and to assess model fidelity in predicting PAL directivity. Ambient conditions were fixed at 20 °C and 50% RH. For quantitative evaluation, we computed RMSE against the measured difference-frequency directivity at 2.0 m (dB re 20 µPa) after peak-normalising each curve to 0 dB at the main-lobe maximum; model outputs were sampled on the same angles from −90° to 90° in 5° steps with no smoothing or weighting. We also report a main-lobe RMSE within ±20° about the beam axis.

Using the measurements as reference, global RMSE values are 45.23 dB for the Convolutional model, 28.40 dB for the Westervelt-based model, and 15.51 dB for the Dual-Convolutional model. Restricting the analysis to the main lobe reduces the errors to 24.59 dB, 12.22 dB, and 6.84 dB, respectively. These results show that main-lobe prediction is consistently more accurate than off-axis behaviour for all methods, and that the Dual-Convolutional model

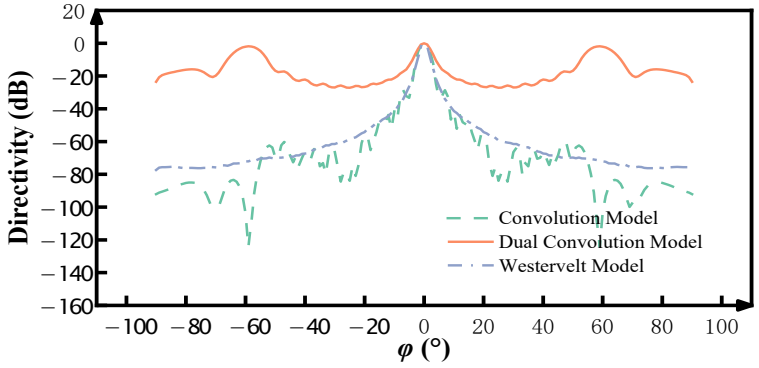
provides the closest match to measurement both globally and within the main lobe, making it the most suitable basis for subsequent design studies.

3.4.2 Primary Waves without Steering Angle

Fig. 3-5 examines the frequency dependence of side lobe formation by fixing the steering at 0 degree and comparing two difference-frequency cases, DF = 8 kHz and DF = 4 kHz. The figure contrasts the predicted directivity of the Dual Convolutional, Convolutional, and Westervelt-based models at the same operating conditions. Across both frequencies, the models yield distinct sidelobe patterns. The Dual Convolution model produces pronounced secondary lobes near the main beam. The Convolution model shows multiple sidelobes, with a higher count at 8 kHz, which agrees with expected frequency scaling. The Westervelt-based result shows more suppressed off-axis energy. These results show that sidelobe behaviour depends on both the difference-frequency and the chosen model.



(a) Difference Frequency = 8000Hz



(b) Difference Frequency = 4000Hz

Fig. 3-5. The result of directivity of three kinds of model methods at DF (a) 8000 Hz and (b) 4000 Hz.

3.4.3 Steered Primary Waves

At a fixed audio difference-frequency of 4 kHz, with a 40 kHz carrier and 10 mm element spacing, we compare directivity at steering angles 0˚,30˚, and 60˚ using the Convolutional, Dual-Convolutional, and Westervelt-based models. Because the spacing exceeds half the carrier wavelength in air, grating lobes are expected. At broadside, the first grating lobe appears near ±60°, which is evident in Fig. 3-6.

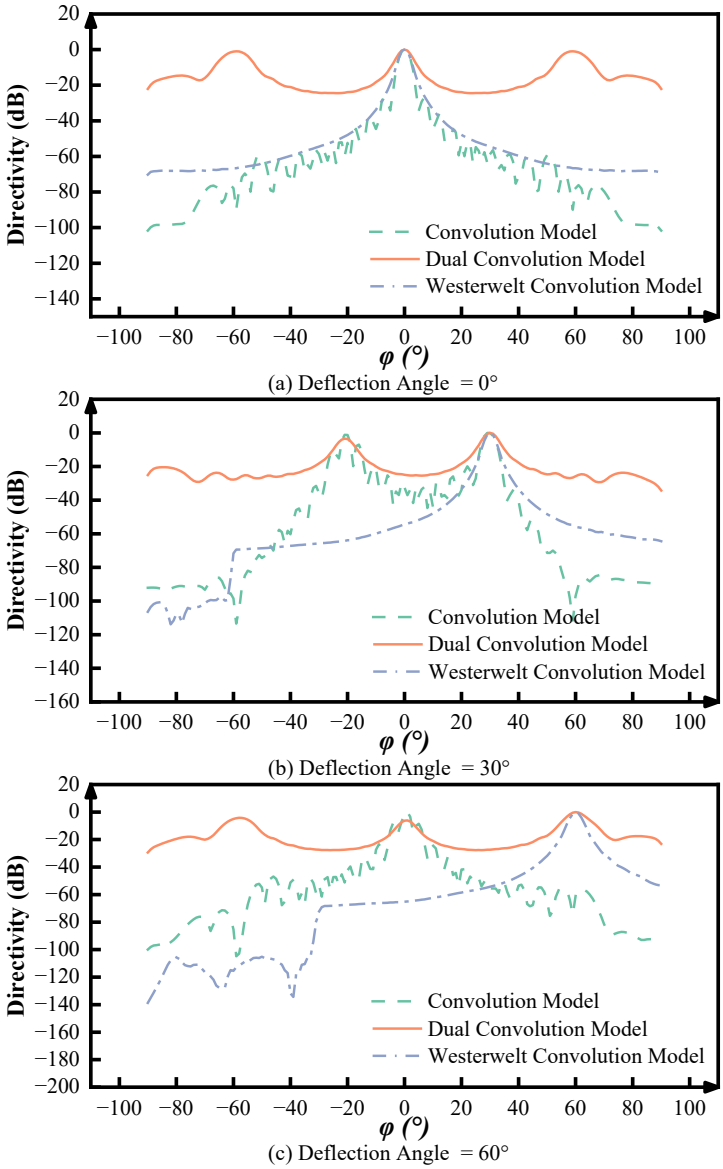


Fig. 3-6. The result of directivity of three kinds of model method, at, steering angle θ= 0°, 30°,60° respectively. (The space between each transducer: 10mm, Carrier wave frequency: 40 kHz, DF: 4 kHz)

As the steering angle increases to 30° and 60°, these lobes move toward the main beam, consistent with standard array behaviour. The Dual-Convolution model reproduces these lobe positions and peak-side lobe levels most faithfully across all three steering cases. The Westervelt-based model under-attenuates off-axis energy (higher PSL), whereas the Convolution model over-attenuates side lobes and slightly broadens the main lobe. All curves are peak-normalised to 0 dB at the main-lobe maximum and plotted on matched angle grids.

These trends have direct design implications. For 40 kHz in air, avoiding visible grating lobes for all steering would require $d \leq \lambda_c/2 \approx 4.3$ mm, which is rarely practical with standard transducers. When $d > \lambda_c/2$, steering pulls aliases toward the main beam, increasing peak-sidelobe level; mitigation then relies on non-uniform spacing and mild apodisation to trade a small increase in main-lobe width for lower sidelobes. Because the audible difference-frequency beam inherits the spatial structure of the ultrasonic primary, these spacing- and steering-dependent behaviours persist after demodulation, motivating the sparse, PSO-designed layouts evaluated in Chapter 4.

3.4.4 Effect of Carrier Frequency

The ultrasonic carrier (primary) frequency f_c governs the primary beam width, steering range, atmospheric absorption, through the sampling ratio d/λ_c with $\lambda_c = c_0/f_c$—susceptibility to grating lobes. We therefore evaluate $f_c \in \{40,60,80\}$ kHz in Fig. 3-7. A carrier of 40 kHz is adopted as the pragmatic baseline: commercial PZT transducers typically resonate near 40 kHz and deliver higher acoustic output for a given drive; atmospheric absorption in air increases with frequency, reducing usable range at 60–80 kHz under comparable conditions; and, for a fixed element pitch d_r, a lower f_c increases λ_c and thus decreases d/λ_c, alleviating grating-lobe formation during beam steering. The 60 kHz and 80 kHz cases are retained to quantify frequency sensitivity.

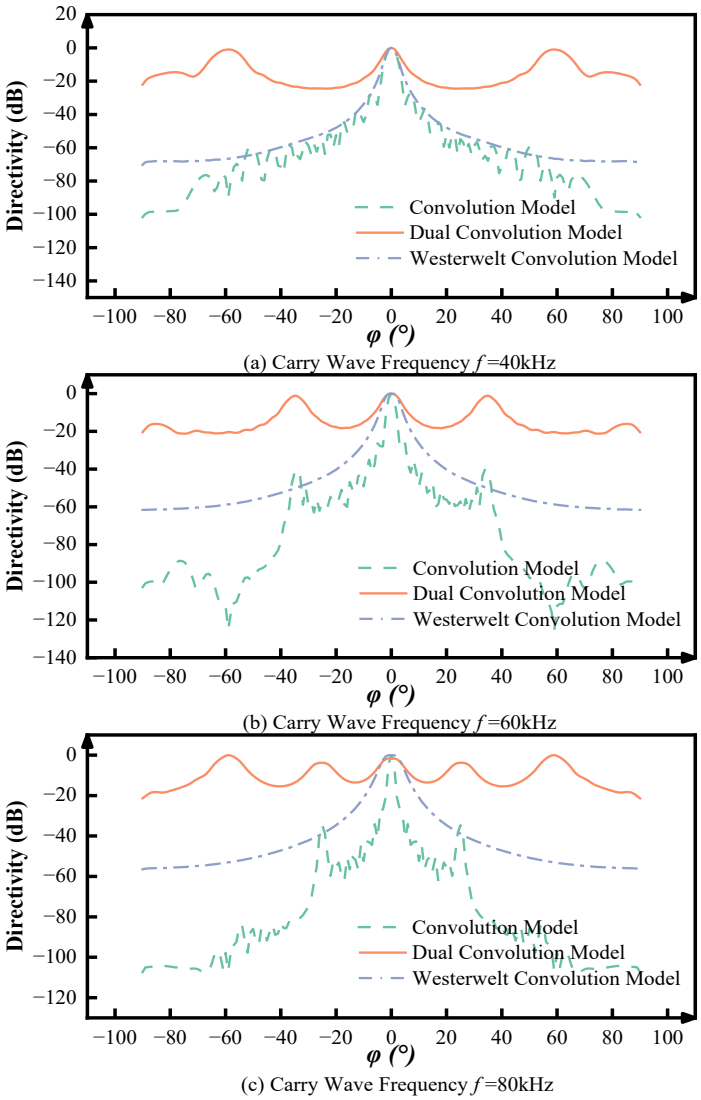


Fig. 3-7. The result of directivity of three kinds of model method, at 40 kHz, 60 kHz and 80 kHz, respectively. (The space between each transducer: 10mm, steering angle: 0°, DF: 4 kHz).

As shown in Fig. 3-7, with the element pitch d fixed, increasing the carrier frequency f shortens the wavelength $\lambda=c_0/f$and reduces the angular spacing between grating lobes, so the pattern appears denser. For a steered uniform array, the grating-lobe angles satisfy $\sin\theta_m \approx \sin\theta_0 + m\,\lambda/d$ ($m\in Z$), giving a spacing in $\sin\theta$ of $\Delta \approx \lambda/d \propto 1/f$; higher f therefore introduces more aliased lobes within the observable angular range. At the same time the main lobe narrows, improving on-axis directivity but increasing susceptibility to grating-lobe energy during steering. Because the audible difference-frequency beam inherits the ultrasonic primary's spatial structure, these frequency-dependent sampling effects persist after demodulation,

revealing a trade-off among directivity, grating-lobe suppression, and propagation loss as f increases.

3.4.5 Effect of Transducer Spacing:

At a carrier of 40 kHz, we model the parametric array loudspeaker as a uniform linear array and compare three practical inter-element spacings: 5 mm, 10 mm (typical of 10 mm-diameter transducers), and 15 mm. As shown in Fig. 3-8, spacing strongly governs both the appearance and the position of grating lobes.

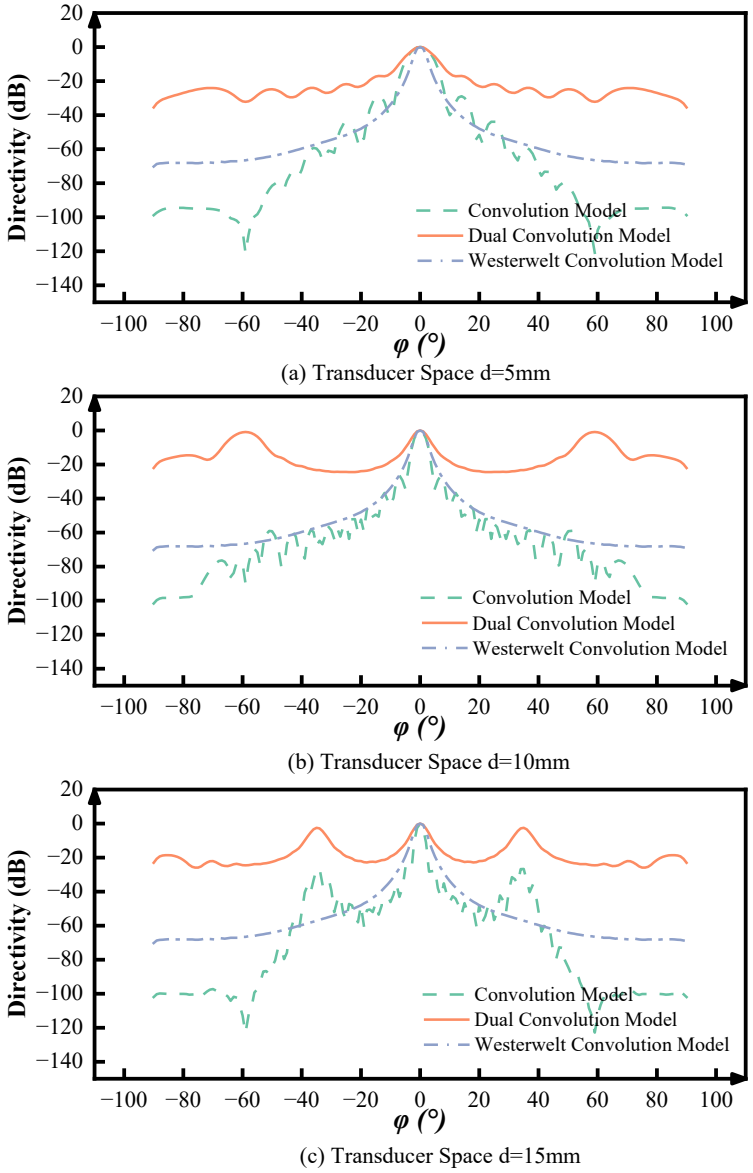


Fig. 3-8. Directivity patterns of three transducer spacings: 5 mm, 10 mm, and 15 mm (carrier: 40 kHz, steering angle: 0°, DF: 4 kHz).

With 5 mm spacing, the pattern remains essentially single-lobed with no visible grating lobe in both of Convolution model and Dual Convolution model. At 10 mm, a prominent pair of grating lobes emerges at large off-axis angles (≈ ±60°). Increasing the spacing to 15 mm shifts these lobes closer to the main beam (≈ ±35°) and elevates their levels, indicating stronger off-axis energy. Relative tendencies are stable: the Westervelt-based curve tends to show a higher off-axis floor and stronger ripple; the simple Convolution model tends to attenuate side lobes more and slightly broaden the main lobe; the Dual-Convolution result tends to sit between these two in both peak side lobe level. These observations highlight that tighter spacing suppresses grating-lobe formation, whereas wider spacing brings lobes toward broadside and increases their amplitude, underscoring the need to set the element pitch carefully when targeting low side lobe levels in PAL array design.

3.5 Summary

This chapter introduced a dual-convolution directivity model for parametric array loudspeakers. The model combines the product of primary patterns with a Westervelt-type directivity term and a phased-array factor. By making aperture spacing explicit, it accounts for grating-lobe formation under steering and improves the interpretability of far-field predictions. Its convolutional form supports fast evaluation on matched angle grids and gives a direct link between array geometry, element spacing, and off-axis energy.

Model fidelity was tested against measurements at 2.0 m using peak-normalised dB directivity from −90° to 90° in 5° steps, with an extra ±20° main-lobe window. Under these settings, the dual-convolution model achieved the lowest errors: global RMSE 15.51 dB versus 28.40 dB for the Westervelt-based model and 45.23 dB for the simple convolution model. Within the main lobe the RMSEs were 6.84 dB, 12.22 dB, and 24.59 dB, respectively. The model also reproduced the grating-lobe directions predicted by d/λ sampling, giving reliable placement of secondary peaks under steering.

However, Traditional analytical methods rely on simplified assumptions. They often ignore nonlinear propagation and array complexity. Conventional numerical methods can also fall short. They may impose rigid geometries or lack sufficient resolution.

To address these limits, we use FEM as the simulation backbone in Chapter 4. FEM models wave interactions in complex spatial domains with high fidelity. It provides the datasets needed to train and test our framework. It also offers a robust reference for validating the proposed predictions under realistic conditions.

.

Chapter 4

PSO-Based Optimisation of Sparse Arrays and Experimental Validation of Far-Field Directivity

To enhance the directional performance of PALs, this chapter presents an integrated framework that combines FEM simulation, particle-swarm optimisation (PSO) for array design, and experimental validation. Unlike traditional analytical models, the FEM approach offers greater flexibility and physical accuracy. It accounts for nonlinear propagation, transducer coupling, and complex boundary conditions. It is therefore well suited to high-resolution prediction of far-field sound fields.

To verify the framework, we run controlled experiments with a custom 8 × 8 PAL array. We test both uniform and PSO-optimised sparse layouts. We then compare the measured directivity with the simulation outputs. The results show strong agreement between prediction and measurement. This confirms the validity of the FEM-based model and the practical benefit of PSO-guided array design.

4.1 Introduction

In acoustic engineering, adaptive PALs advance traditional designs by using phased-array control to steer audio beams without mechanical rotation. The approach is well studied and deployed in ANC and sound reproduction. An adaptive PAL typically comprises a linear array of ultrasonic transducers, utilizing phased array technology to control beam directionality [104].

To avoid grating lobes, element spacing is set to half the wavelength of the ultrasonic carrier wave($\lambda_c=c_0/f_c$), i.e., $d\leq\lambda_c/2$. Using the highest operating frequency ensures lobe-free performance across the steering band; for a uniform linear array scanned to $\theta_{\max}$, a common guideline is $d/\lambda_c\leq 1/(1+\sin\theta_{\max})$. In practice, meeting this spacing is difficult because commercial ultrasonic transducers have centimetre-scale apertures and the frequency is usually 40 kHz, so arrays often exceed the ideal density and rely on weighting and optimisation to

suppress residual grating energy [105]. To mitigate side lobe interference and spatial aliasing when the spatial-Nyquist condition cannot be met across the steering band, we adopt a sparse non-uniform array as shown in Fig. 4-1.

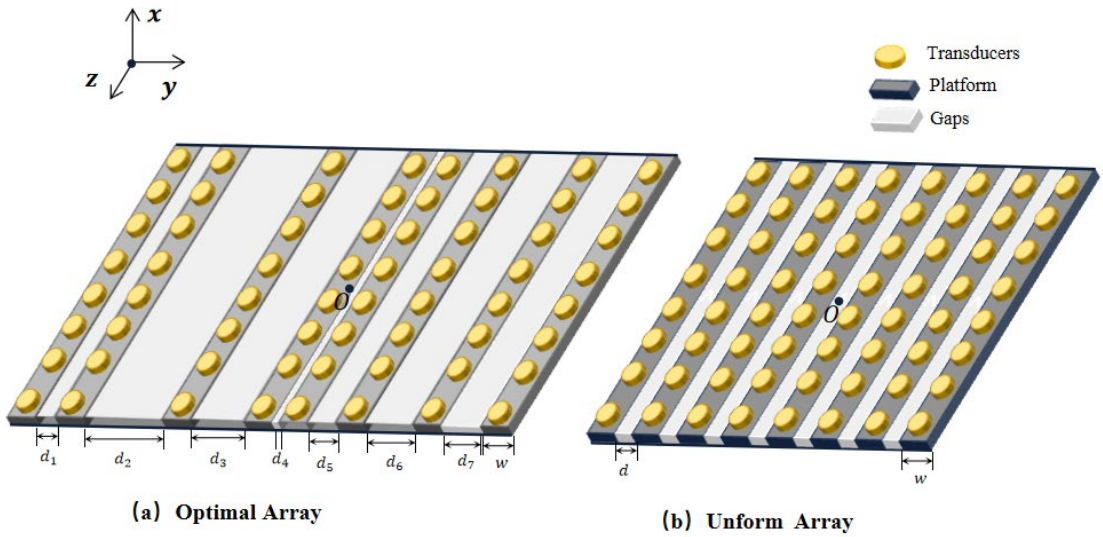


Fig. 4-1. Schematic representation of (a) optimal array and (b) uniform array of the PALs shows *N* ultrasonic transducers (represented as 8 sets).

In the 2D formulation, only the inter-column spacings along x are optimised, while the within-column pitch along y is fixed at $d_y=\lambda_c/2$(set by the highest carrier) for manufacturability and to avoid aliasing in y. Because steering and evaluation are performed in the x–z (elevation) plane, the rectangular aperture factor separates; with uniform $d_y$the y-factor is effectively constant, so side lobe control reduces to shaping the x-aperture. We therefore pose a constrained optimisation over the inter-column centres $\{d_1,\ldots,d_7\}$(optionally with amplitude weights), subject to overall aperture-width $w$and spacing/order bounds, and solve it via global search with FEM-based forward evaluation. Fig. 4-1 summarises the geometry and baseline comparison: the array-fixed frame (x,y,z) is centred at Oon the baffle plane $(z=0)$; eight columns of identical transducers are aligned with y; the uniform baseline uses a constant inter-column spacing d along x, whereas the optimised layout employs aperiodic spacings $\{d_1,\ldots,d_7\}$(with $d_y$as above), which mitigates periodic grating-lobe reinforcement within the target steering band and reduces the peak side lobe level.

Because the sparse layout introduces a large, nonconvex design space without a closed-form solution, we employ a PSO-based global optimisation coupled with FEM evaluation. The

optimizer searches over the inter-column spacings ($d_1,\ldots,d_7$) (equivalently, the column centres referenced to **O**) to minimize the penalized side lobe objective defined in Eq. (4-16), while respecting the spacing bounds and maintaining the aperture centred at **O**. This PSOFEM procedure yields transducer positions that achieve high main-lobe directivity with reduced peak side lobe level at the specified steering angles.

Additionally, to accurately model the acoustic field and evaluate the impact of different array configurations, we employ the FEM, which provides a high-fidelity simulation of nonlinear acoustic wave propagation to generator the Far-field data. Traditional analytical approaches, such as the Westervelt equation and the KZK equation, offer useful approximations but are often limited by simplifying assumptions, such as paraxial approximations or restricted boundary conditions. FEM, in contrast, provides greater flexibility in handling complex boundary conditions, irregular geometries, and near-field effects, making it particularly effective for modelling the intricate interactions within PALs. By resolving the nonlinear effects more accurately, FEM enables a deeper understanding of wave behaviour, diffraction, and beamforming characteristics, ensuring a more realistic representation of sound propagation in practical applications [106], [107].

By integrating PSO-based 2D array optimisation with FEM-based acoustic field simulations, this study presents a comprehensive and computationally efficient framework for improving PAL performance and build an experiment setup to support the result. This combination ensures precise beamforming control, reduced side lobe interference, and enhanced adaptability, making it well-suited for real-world directional audio systems. The proposed approach not only advances theoretical modelling but also offers practical benefits in applications such as public address systems, spatial audio design, and automotive acoustics, where accurate sound projection is crucial.

4.2 PSO-Based Optimisation of Sparse Array Configuration for Far-Field Directivity Enhancement

This section introduces a hybrid optimisation framework that combines FEM simulations and PSO to improve the far-field directivity of PALs through sparse array design. We first present the COMSOL-based modelling of the parametric loudspeaker in both 2D and 3D, then justify

our choice to proceed with 2D for the main study: at ultrasonic resolution, the 3D FEM problem increases Degree of freedom (DOF) by 1 order of magnitude and demands sub-millimetre meshing in the domain and PML, making parameter sweeps and optimisation intractable within our compute budget. Using the validated 2D model, we conduct array optimisation using particle swarm optimisation with grid-search baselines, targeting side lobe suppression and reduced RMSE of the predicted directivity over prescribed steering angles. PSO-Based Optimisation of Sparse Array Configuration for Far-Field Directivity Enhancement

4.2.1 2D Far-Field Directivity Calculation By FEM

In the study, we utilize the Pressure Acoustics Frequency Domain interface from the Pressure Acoustics branch for modelling acoustic wave propagation in fluids. This is applied to calculate pressure variations for acoustic wave propagation in fluids. It is ideal for frequency-domain simulations involving harmonic pressure field variations, which form the core of the model used in this thesis. To optimize computational resources, our model is constructed in two dimensions, incorporating a line sound source which produces a cylindrical sound wave. The physics interface solves the Helmholtz equation in the frequency domain for specified frequencies or performs an eigenfrequency or mode analysis study. This interface efficiently computes pressure variations by solving the Helmholtz equation in the frequency domain and is particularly suited for frequency-domain simulations involving harmonic pressure field variations. The primary variable, the sound pressure, p_t, encapsulates the acoustic perturbations relative to the ambient pressure. For 2D models, the default out-of-plane wavenumber k_z is set to 0 rad/m, indicating no variation along the z-axis, thereby simplifying the pressure expression in our simulations.

Fig. 3-1 (Chapter 3) is a schematic that explains the PAL architecture and self-demodulation mechanism, including the formation of a highly directional main lobe and the presence of side lobes. To facilitate computation, Fig. 4-2 depicts the Pressure Acoustics (Frequency Domain) model used in this study. The computational region is a semicircle of radius 200 cm centred on the array and terminated by a perfectly matched layer (PML) to suppress boundary reflections. A mesh inset is provided; the minimum element size is $\lambda_{min}/6$ to resolve the highest frequency. Far-field responses are sampled on an observation arc at 200 cm, matching the experimental microphone distance. This choice places the sampling arc well beyond the array's near-field

(Rayleigh) region at 40 kHz and 42 kHz, making the angular pattern effectively range-invariant and enabling a direct simulation-to-measurement comparison.

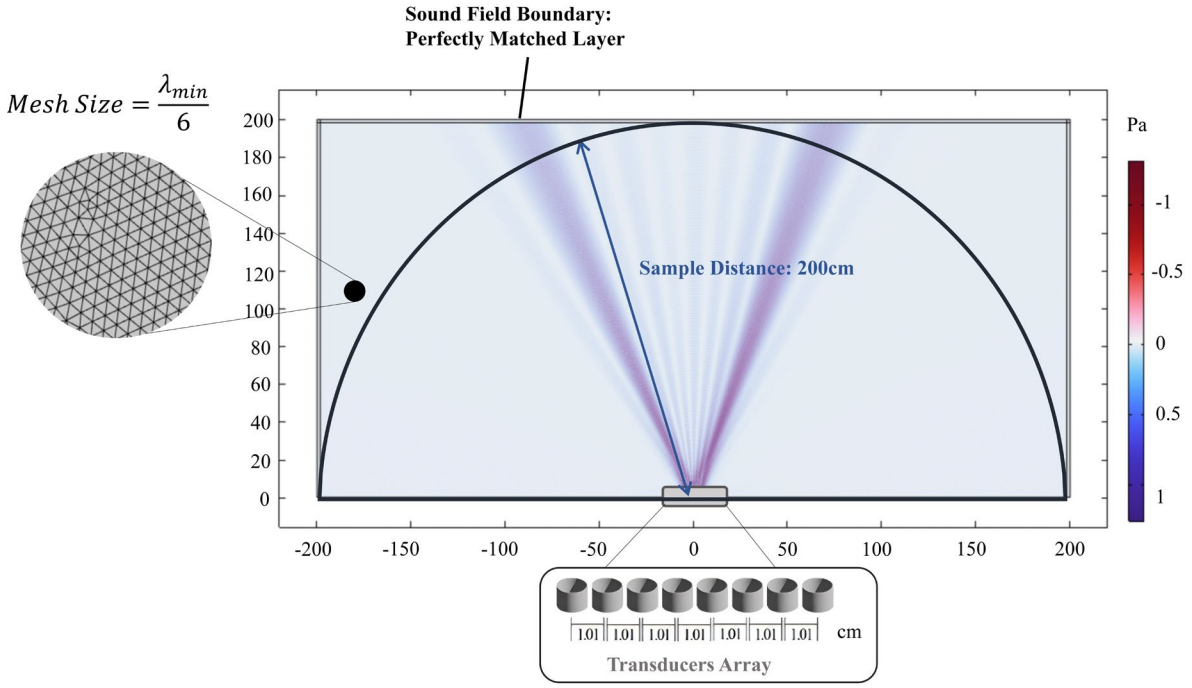


Fig. 4-2. 2D Acoustic Field Simulation Model in COMSOL with Perfectly Matched Layer (PML) Boundary.

We compute the difference-frequency far-field in a 2D axisymmetric pressure-acoustics model consistent with the principal *y-z* evaluation plane. Unless stated otherwise, the DF is 2 kHz; air properties follow the built-in model at 21˚C,1 atm,RH≈50%. For Fig. 4-2 we use a uniform array with 8 elements and 1.01 cm spacing. All training and evaluation were run on a single NVIDIA RTX 3080 Ti (12 GB VRAM). No distributed or multi-GPU training was used and identical preprocessing/normalization/evaluation code was applied across all runs.

This is achieved by using the line source on axis node to introduce a line source along the axis of symmetry in 2D axisymmetric components. This line source represents a radially vibrating cylinder in the limit as its radius approaches zero. The sound pressure value of any point in the calculation space can be obtained by solving the partial differential equation (4-1) by the FEM [80], [108].

$$\nabla\cdot\left(-\frac{1}{\rho_c}\left(\nabla p_{\mathrm{t}}-\mathbf{q}_{\mathrm{d}}\right)\right)-\frac{k_{eq}^2 p_{\mathrm{t}}}{\rho_c}=\frac{4\pi}{\rho_c}S\delta(z-z_0)dz \tag{4-1}$$

$$k_{eq}^2=\left(\frac{\omega}{c_c}-i\alpha\right)^2-k_z^2 \tag{4-2}$$

Where p_{t} represents the total acoustic pressure, ρ is the fluid density, c is the speed of sound, $\mathbf{q}_{\mathrm{d}}$ denotes the dipole domain source, complex density $\rho_c=\frac{\rho c^2}{c_c^2}$, complex speed of sound as $c_c=\omega/k-i$. In this context, $\delta(z-z_0)$ is the 2D delta function that introduces the source at the axis of symmetry, where $z=z_0$ and $r=0$, while dz represents the line element along the z axis (SI unit: m). The monopole amplitude S (SI unit: N/m^2) is determined by the type of source selected and is analogous to the 3D case for a line source. To include an edge source at $r=r_0$, define it using the volume flow rate per unit length Q_S and the phase φ of the source. The flow edge source then establishes the monopole amplitude as (4-3):

$$S=e^{i\varphi}\frac{i\omega\rho_c}{4\pi}Q_{\mathrm{S}} \tag{4-3}$$

A flow edge source with strength Q_S represents an area flow emitted from the source, modelled as a very thin cylinder with a pulsating surface. Specify Q_S (SI unit: m^2/s), the volume flow rate per unit length from the source, as the amplitude for the source strength in the field.

$$p_t=p+p_b \tag{4-4}$$

In the frequency domain (using Fourier transforms), the nonlinear correction becomes[108]:

$$\hat{p}_b(\omega)=\frac{\beta}{\rho c^4}(-\omega^2)\hat{p}^2(\omega)-\frac{i\delta\omega}{\rho c^2}\nabla^2\hat{p}(\omega) \tag{4-5}$$

Where $\hat{p}(\omega)$ represents the Fourier transform of the primary pressure p, $\hat{p}_b(\omega)$ denotes the Fourier transform of the nonlinear correction p_b. This expression highlights that the nonlinear component p_b scales with the square of the primary pressure amplitude. As a result, higher frequencies and larger pressure amplitudes lead to more pronounced nonlinear effects. This

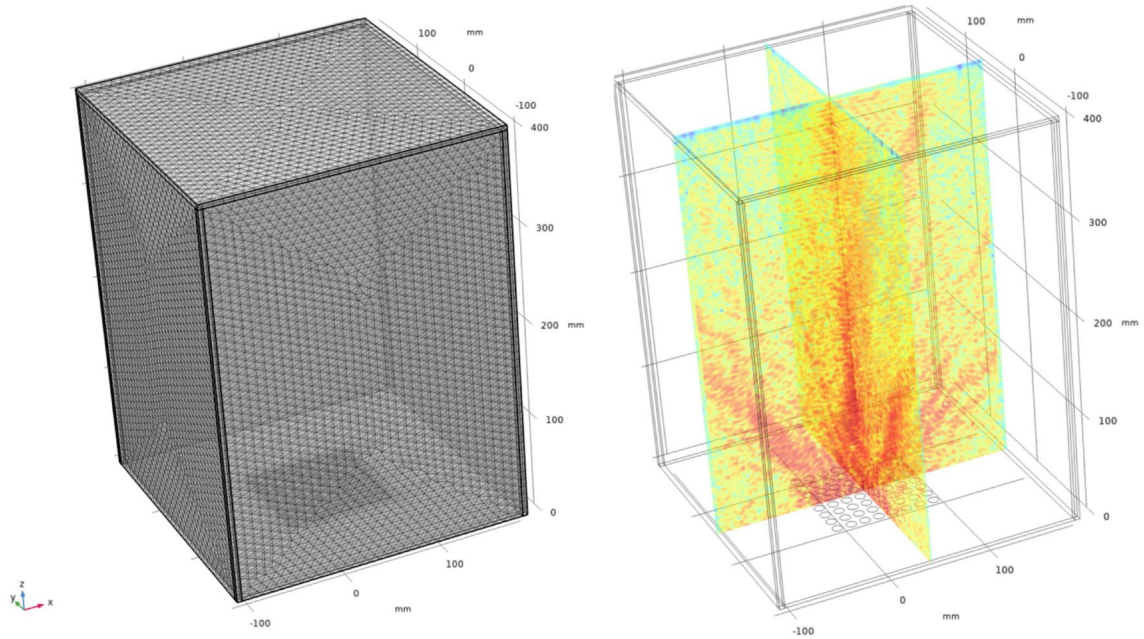
formulation plays a crucial role in understanding the generation of secondary sound fields in PALs, particularly in predicting far-field directivity patterns. To simplify the analysis, we only consider the second harmonic, which is the most significant nonlinear component. Under this assumption, the nonlinear correction can be approximated as:

$$p_b \approx \frac{\beta \omega^2 p^2}{\rho c^4} \tag{4-6}$$

4.2.2 3D Far-Field Model By FEM

This section clarifies why we do not adopt full 3D FEM for routine directivity simulations: at ultrasonic resolution it is computationally prohibitive in both degrees of freedom and runtime. For context only, Fig 4-3 is presented as an illustrative 3D FEM model. Fig 4-3 (a) is the mesh of the model and the Fig 4-3 (b) are the Pressure-magnitude slices $|p_t|$ on two orthogonal planes for a representative case (40 kHz); The example uses an 8×8 uniform array with triangular size of 10 mm over a 300 × 300 × 400 mm rectangular domain.

Moving from 2D to 3D increases the number of finite-element unknowns by roughly one to two orders of magnitude at comparable resolution: DOF$\propto(L/\lambda)^3$ (3D) versus $(L/\lambda)^2$ (2D). At ultrasonic frequencies (40-42 kHz in air), a conservative mesh of $\lambda_{\min}/6$ implies sub-millimetre elements in both the physical domain and the absorbing layer, which leads to substantially higher memory usage, longer factorization/iteration times for the complex Helmholtz system, and tighter requirements on preconditioning and layer thickness. When weak nonlinearity is included via the frequency-domain correction p_b, the volumetric coupling further amplifies the cost.



(a) (b)

Fig. 4-3. 3D COMSOL model. (a) Finite-element tetrahedral mesh containing the 8×8 uniform array with triangular size of 10 mm over a 300 × 300 × 400 mm rectangular domain; (b) Pressure-magnitude slices $|p_t|$ on two orthogonal planes for a representative case (40 kHz , steered to ψ).

Even with this coarse discretisation, included only to visualise the geometry and field slices, our hardware barely completes the solve. The study focuses on principal-plane far-field directivity of a slender, effectively linear aperture and reports relative SPL metrics. Under these conditions, a 2D surrogate captures the dominant beam physics, including main-lobe formation, steering, and side lobe behaviour, because out-of-plane variations are weak over the apertures and steering angles considered. Quantitative assessment is performed on 2D slices in the principal planes, as in Fig 4-3 (b), so a full 3D model would provide the same analysis views at substantially higher computational cost. Enforcing the experimental parameters in three dimensions, such as a 200 cm propagation distance and a $\lambda_{min}/6$ mesh size, is impractical on the available hardware. For efficiency and clarity, we therefore retain the validated 2D formulation and use the 3D model only for illustration and scoping.

4.2.3 Theory and Implementation of PSO

Many metaheuristic algorithms have been used to optimise element placement in phased-array antennas. Examples include Teaching–Learning-Based Optimisation (TLBO) [109], modified Quantum Particle Swarm Optimisation (QPSO) [110], Symbiotic Organisms Search (SOS) [111], and machine learning-enhanced genetic algorithms [112] have been employed to generate tailored beam patterns, such as pencil beams or flat-top responses, while minimizing side lobe levels [113]. They frame beamforming as a nonlinear, high-dimensional optimisation problem. The same methodology transfers to parametric speaker arrays and supports configuration design and directivity control.

Among the many optimisation algorithms, PSO, SA (Simulated Annealing), and GA(Genetic Algorithm) are widely applied in high-dimensional design problems, including phased array optimisation. Each method presents unique strengths and limitations:

- **PSO** is favoured for its rapid convergence and ease of parallelization, making it suitable for large-scale problems. However, it is sensitive to parameter selection and prone to premature convergence if not properly tuned [114].
- **SA** offers robust global search capabilities by probabilistically accepting worse solutions in early iterations, helping to escape local minima. Its efficiency, however, is constrained by the need for careful temperature schedule design and its inherently sequential nature [113].
- **GA** maintains a diverse population of candidate solutions, reducing the likelihood of local convergence. Nonetheless, it often exhibits slower convergence rates and requires extensive parameter tuning to balance exploration and exploitation [115], [116].

These comparisons guide the choice of optimisation algorithms under specific application needs and compute limits. In this study, we choose PSO for array configuration because it balances simplicity, scalability, and proven effectiveness.

As summarized in Table 4-1, each algorithm exhibits distinct strengths and limitations: PSO converges quickly and parallelises well, but it is sensitive to parameter tuning and can converge prematurely. SA escapes local optima by accepting worse solutions early on, but it converges slowly and depends strongly on the temperature schedule. GA maintains population diversity and reduces the risk of local minima, but it often converges more slowly and needs complex

parameter control in high-dimensional cases. Given these trade-offs, PSO is a practical and computationally efficient choice for the optimisation tasks in this work.

Table 4-1. Comparative Analysis of Optimisation Algorithms: PSO, SA, GA.

Algorithm	Reference	Advantages	Limitations
PSO	[114], [117]	Simple & parallel; fast on continuous spaces;	Premature convergence risk; Sensitive to parameters
SA	[113], [118]	Escapes local optima; simple; convergence under proper cooling.	Slow in practice; schedule-sensitive; weak parallel efficiency.
GA	[115], [119]	Maintains diversity; handles discrete/complex encodings.	Slower convergence; many hyperparameters; diversity control needed.

PSO is an evolutionary algorithm inspired by the social behaviours of animal groups such as bird flocks and fish schools. Proposed by Kennedy and Eberhart in 1995, PSO has proven effective for addressing nonlinear and multi-dimensional optimisation problems. The algorithm's simplicity and robustness have contributed to its widespread adoption across diverse domains.

In PSO, each particle's position is a D-dimensional vector that represents a candidate solution. We evaluate each position with an objective function. The goal is to let all particles explore the space and move toward the global best solution. Each particle stores its own best position (personal best). It also knows the best position found by the swarm. This personal–social update lets particles adjust their motion using both private and shared information. It improves the search for optima in complex spaces [114]. The position x_i and velocity v_i of each particle i in a D dimensional space is updated at each step of the algorithm according to the following equations:

$$x_i=\{x_{i1},x_{i2},x_{i3},\ldots,x_{iD}\},i=1,2,3,\ldots,N \tag{4-7}$$

The fitness value is calculated by substituting it into the fitness function, and the location of the optimal fitness value experienced by the first particle is noted as the individual historical optimum:

$$p_{besti}=\{p_{\text{besti1}},p_{\text{besti2}},p_{\text{besti3}},\ldots,p_{\text{bestiD}}\},i=1,2,\ldots N \tag{4-8}$$

The optimal position experienced by the entire particle swarm is denoted as the global optimum:

$$g_{\text{besti}}=\{g_{\text{besti 1}},g_{\text{besti 2}},g_{\text{besti3}},\ldots,g_{\text{bestiD}}\},i=1,2,\ldots N \tag{4-9}$$

The velocity of the first particle is given as:

$$v_{ij}=\{v_{i1},v_{i2},v_{i3},\ldots,v_{iD}\},i=1,2,\ldots N \tag{4-10}$$

Then the velocity of the first particle in dimension after the first iteration is.

$$v_{ij}(t+1)=v_{ij}(t)+c_1r_1\left(P_{\text{besti}}(t)-x_{ij}(t)\right)+c_2r_2\left(g_{\text{best}}-x_{ij}(t)\right) \tag{4-11}$$

Where $i=1,2,\ldots N$,$j=1,2,\ldots,D$and r_1,r_2 is the random number. Location updated for:

$$x_{ij}(t+1)=x_{ij}(t)+v_{ij}(t+1) \tag{4-12}$$

where $i=1,2,\ldots N$,$j=1,2,\ldots D$.

In PSO, the calculation time primarily depends on the number of particles N and iterations ger. The maximum function evaluations are given by:

$$Number\ of\ Calls=N\times ger \tag{4-13}$$

The overall workflow of the PSO algorithm is illustrated in Fig. 4-4. The process begins with the random initialization of a population of particles, each representing a candidate solution within the search space. At each iteration, particles adjust their velocities and positions based on two key influences: their personal best-known positions p_{besti} and the global best position g_{besti} discovered by the swarm. This balance between individual learning and group knowledge supports both exploration and exploitation. The loop continues until a stopping rule is met, such as a maximum number of iterations or a target fitness. Through this process, PSO finds near-optimal solutions in complex, high-dimensional

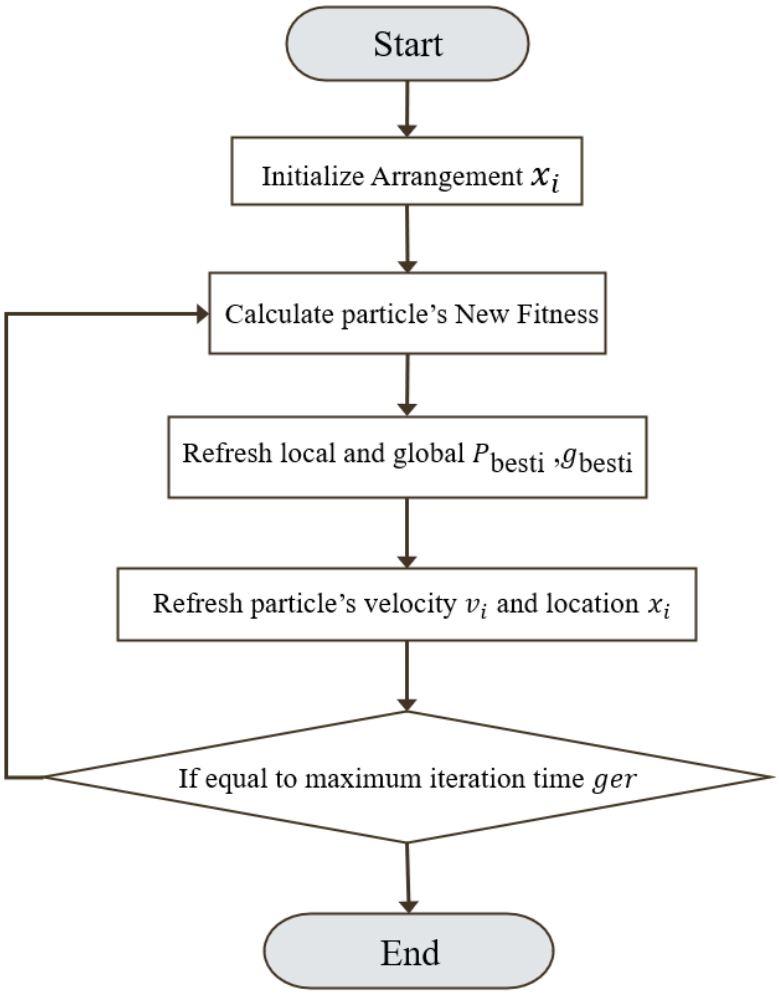


Fig. 4-4. Detailed Flowchart of the PSO Algorithm for Optimizing Speaker Array Configurations.

To study the effect of sparse arrays on performance, we normalised the initial SPL to 1 dB and varied the inter-element spacing from 0 to 13 mm. The upper bound of 13 mm was chosen to ensure that the total array length remained within twice the acoustic wavelength, thereby avoiding grating lobes associated with oversized apertures. The cost function and sampling strategy followed the methodology described in chapter 4.2.4, with all array configurations centred at the origin. Environmental conditions were fixed across all trials, with air humidity set at 75% and ambient temperature at 13°C, to ensure consistency in acoustic propagation parameters.

4.2.4 Cost Function Design and Algorithmic Implementation

In this study, the cost function equals the SPL of the main lobe minus the SPL of the highest sidelobe. We add a constraint to keep the main lobe above a minimum level. To explore the

effects of sparse array configurations, the initial SPL is normalized to 1, and the transducer spacing varies from 0 cm to 13 mm. The upper limit is selected to keep the overall array aperture within twice the acoustic wavelength, thereby avoiding excessive array size and potential spatial aliasing. The sampling strategy and cost function evaluation follow the methodology described previously, with all arrays symmetrically centred at the origin.

To prevent the main lobe from becoming too narrow or weak, an additional constraint is introduced based on the SPL of the main lobe, denoted as L_{Main} defined as equation (4-14):

$$J_{\text{tot}}(x)=\underbrace{\left(\max_{k} L_{\text{Side}}^{(k)} - L_{\text{Main}}(\psi)\right)}_{J(x)} + \lambda \max(0, L_{\min} - L_{\text{Main}}(\psi)) \tag{4-14}$$

Here, ψ denotes the target steering angle. The main-lobe window $\Delta\varphi_{\text{main}}>0$ is chosen to cover the main lobe around ψ; the corresponding peak level is $L_{\text{Main}}(\psi)=\max_{\varphi\in[\psi-\Delta\varphi_{\text{main}},\psi+\Delta\varphi_{\text{main}}]} L(\varphi)$. The peak side lobe level (PSL) L_{PSL} is computed as the maximum of $L(\varphi)$ outside an exclusion window $[\psi-\Delta\varphi_{\text{excl}},\psi+\Delta\varphi_{\text{excl}}]$ with $\Delta\varphi_{\text{excl}}\geq\Delta\varphi_{\text{main}}$. The design vector $\mathbf{x}=[x_1,x_2,\ldots,x_N]^{\text{T}}$ collects the element centre positions on the array axis (cm), where $x_N-x_{N-1}\in[0,1.3]\ cm$. $L_{\min}$=80 dB and λ=1 unless otherwise stated. PSL serves as the key metric for array performance. The solution $\mathbf{x}_{\text{opt}}$ (optimal element positions) is obtained using the PSO-FEM procedure summarized in Algorithm 1.

Algorithm 1: PSO-Driven FEM Optimisation

Inputs:

ψ # target steering angle (deg)

φ # angle grid for evaluation (deg)

$P_{\text{ref}}=20\mu$ Pa # SPL reference

spacing bounds: [0, 1.3] cm

START

Initialize particle population $\mathbf{x}=[x_1,x_2,\ldots,x_N]^T$

FOR each particle $\mathbf{x}_i$ DO

Transfer $\mathbf{x}_i$ to COMSOL FEM model

Run FEM: compute pressure field $p(\mathbf{r})$

Convert to SPL: $L=20\log_{10}\left(\frac{p_t}{P_{\text{ref}}}\right)$,

Extract directivity pattern $L(\varphi)$

Find main lobe: $L_{\text{Main}}(\psi)=\max_{\varphi\in\Delta\varphi}L(\varphi)$

Find 3 highest side lobes: $\left\{L_{\text{Side}}^{(k)}\right\}_{k=1}^{3}$

Evaluate cost:

$$J_{\text{tot}}(\psi)=\underbrace{\left(\max_{k}L_{\text{Side}}^{(k)}-L_{\text{Main}}(\psi)\right)}_{J(x)}+\lambda\,\max(0,L_{\min}-L_{\text{Main}}(\psi))$$

ENDFOR

Update particles:

$\mathbf{v}_i(t+1)=\omega\mathbf{v}_i(t)+c_1r_1(\mathbf{p}_i-\mathbf{x}_i)+c_2r_2(\mathbf{g}-\mathbf{x}_i)$

$\mathbf{x}_i(t+1)=\mathbf{x}_i(t)+\mathbf{v}_i(t+1)$

ENDFOR

Return optimal positions $\mathbf{x}_{\text{opt}}$, optimal directivity pattern $L(\varphi;\mathbf{x}_{\text{opt}})$

4.2.5 FEM–MATLAB Co-Simulation and Sampling Strategy for Far-Field Directivity Evaluation

In this study, we use COMSOL Multiphysics to implement a finite-element model of the ultrasonic field from a transducer array. Using this method, detailed acoustic pressure field distributions are calculated, providing comprehensive insight into how sound waves propagate, interact, and reflect within the defined domain. As shown in the Integration Diagram (Fig. 4-5), these field data feed the optimisation stage. They allow precise evaluation of directivity and sidelobe suppression, which are key metrics for optimising the array..

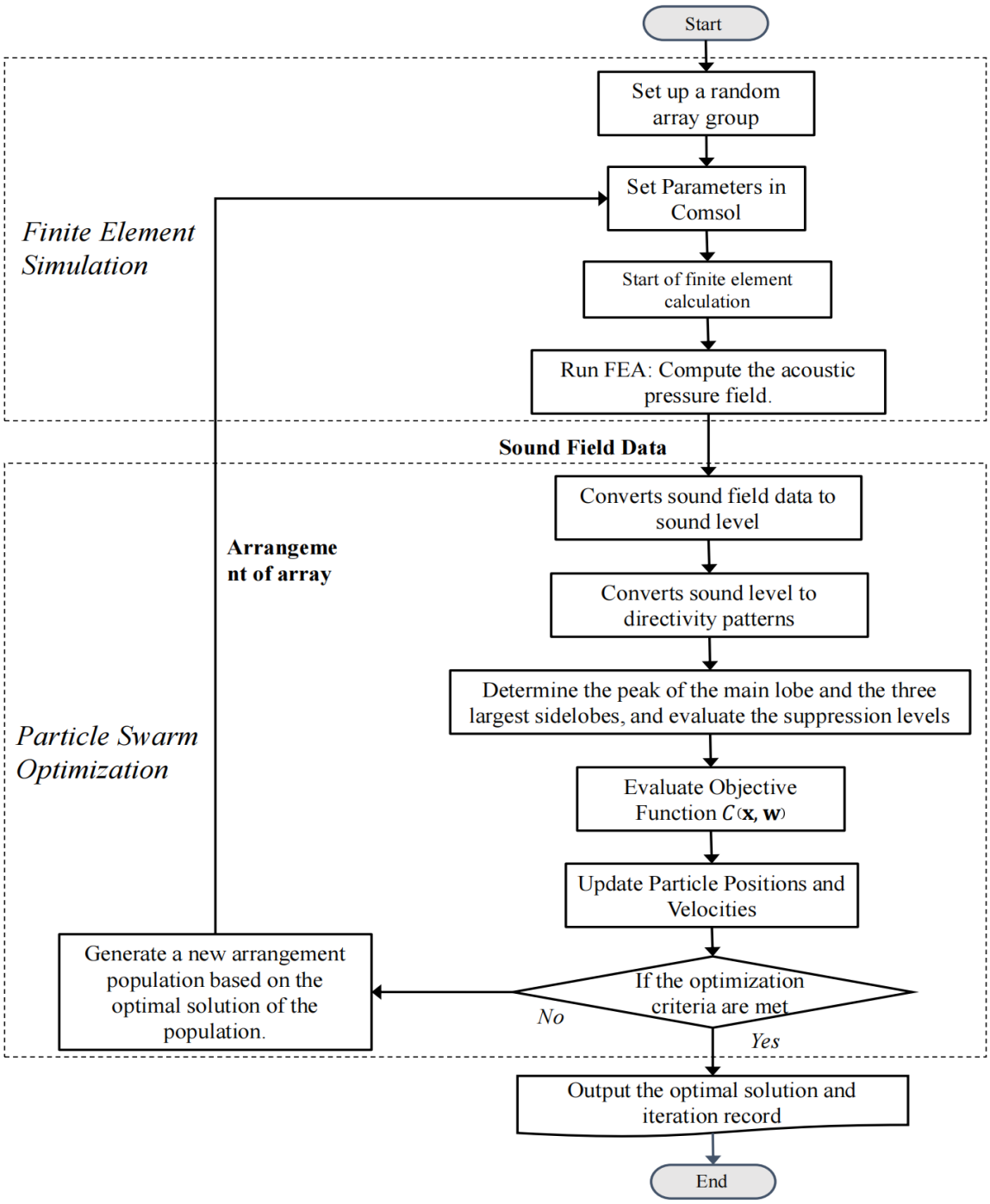


Fig. 4-5. Integration Diagram of COMSOL and MATLAB for Co-Simulation in Speaker Array Optimisation.

As illustrated in the co-simulation workflow (Fig. 4-5), we use the SPL data to extract key performance indicators. These include far-field directivity and sidelobe behaviour. The directivity plots mark the main-beam axis and reveal sidelobes, which indicate off-axis energy

that should be minimised for optimal beamforming. The resulting SPL fields then feed the optimisation stage, so we can evaluate array performance objectively under different layouts.

The simulation environment is run in a rectangular domain, with boundary conditions specifically designed to replicate the experimental setup, thereby minimizing unwanted reflections and refractions that could distort the results. The source is an 8 × 8 array of circular transducers, each 10 mm in diameter, arranged on the bottom plane. The spatial positions of these transducers are dynamically determined by the PSO algorithm implemented in MATLAB. This integrated approach allows for an iterative process, where the PSO algorithm adjusts the array configuration to optimize performance, and the updated parameters are re-applied in COMSOL for further acoustic field calculations.

To ensure accuracy, the medium for the simulation is modelled as air at standard atmospheric pressure, with specific environmental conditions set at 50% humidity and a temperature of 13 ℃. The study focuses on linear acoustic fields, as nonlinear acoustics are excluded; this simplification is necessary because the nonlinear models cannot adequately represent the demodulated sound pressure in this context. MATLAB is used to generate eight-line segments, each 1 cm in length, representing a PAL in 2D. This setup provides a simplified but effective representation of the acoustic field, facilitating detailed analysis and optimisation of the performance of array.

The corresponding simulation results are presented in Section 4.4, where they are compared with experimental measurements to evaluate the accuracy and practical relevance of the modelling approach.

4.3 Experimental Setup and Measurement Protocol for Far-Field Directivity

4.3.1 Experimental Objectives for Far-Field Acoustic Performance

The purpose of this chapter is to experimentally validate the theoretical modelling framework, and the array optimisation strategies presented. In particular, the experiments are designed to assess the far-field acoustic directivity of PALs under both uniform and sparse transducer configurations. These measurements provide critical insight into the practical feasibility and

robustness of the optimisation scheme based on PSO. The experimental campaign is structured around the following key objectives:

- **Characterization of Far-Field Directivity:** To evaluate the spatial distribution of the acoustic field generated by the PAL, with particular focus on quantifying the main lobe width, side lobe level, and the angular suppression of grating lobes. These parameters are critical in assessing the effectiveness of beam control mechanisms in practical directional sound applications.

- **Performance Assessment of sparse arrays:** To determine the degree to which sparse transducer spacing, optimized through PSO, improves the directivity performance relative to conventional uniform arrays. This involves both simulation-based design and hardware implementation of optimized configurations, followed by direct experimental comparison.

- **Analysis of Angular Robustness:** To investigate how the directivity characteristics of the PAL vary with respect to steering angle. Steering angles of 60° and 75° are selected as representative test cases based on prior simulation studies, enabling the analysis of angular performance degradation or stability in both uniform and sparse configurations.

To meet the steering objective, we use the physical setup in Fig. 4-6, The figure shows the array geometry, the beam analysis axes, and the primary/secondary wave paths. In this study, we suppress the secondary wave to simplify control of the phased array. We therefore measure only the ultrasonic primary field at 40 kHz and 42 kHz along a sampling arc orthogonal to the array plane. The microphones scan these angles to reconstruct the primary-field pattern. We then compute the Westervelt directivity from the measured primary field to estimate the corresponding difference-frequency directivity.

Each measurement was conducted across a range of steering angles (e.g., 0°, 60°, and 75°) and transducer configurations (uniform vs. optimized sparse array). At each configuration, the array was rotated in 2.5° increments to capture high-resolution polar sound field data. The measured spatial pressure distributions were subsequently compared to the simulated results for model validation and evaluation of the optimisation's practical effectiveness.

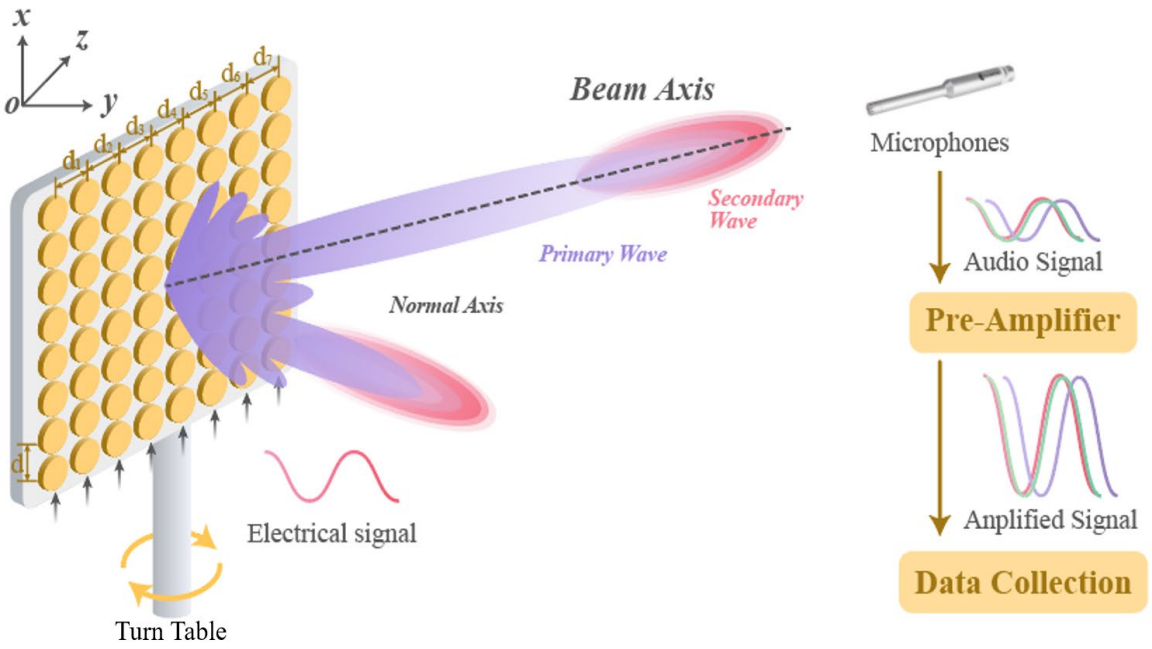


Fig. 4-6. Geometric representation of the PAL beamforming measurement setup.

4.3.2 System Architecture and Hardware Components

All experimental measurements were conducted in a carefully controlled acoustic environment to ensure consistency, suppress reflections, and eliminate external noise. The measurement site was a full anechoic chamber with dimensions 3 m × 3.5 m as shown in Fig. 4-7, located in a separate, acoustically isolated building. The chamber is fully lined with wedge-shaped absorbing foam on all six surfaces, including a suspended mesh floor that allows for free-field conditions while supporting equipment placement. This design ensures reflection-free propagation of acoustic waves across the relevant ultrasonic and audible frequency bands. A temperature-controlled system maintains ambient conditions at 13°C during most measurement sessions, minimizing variations in sound speed due to thermal drift.

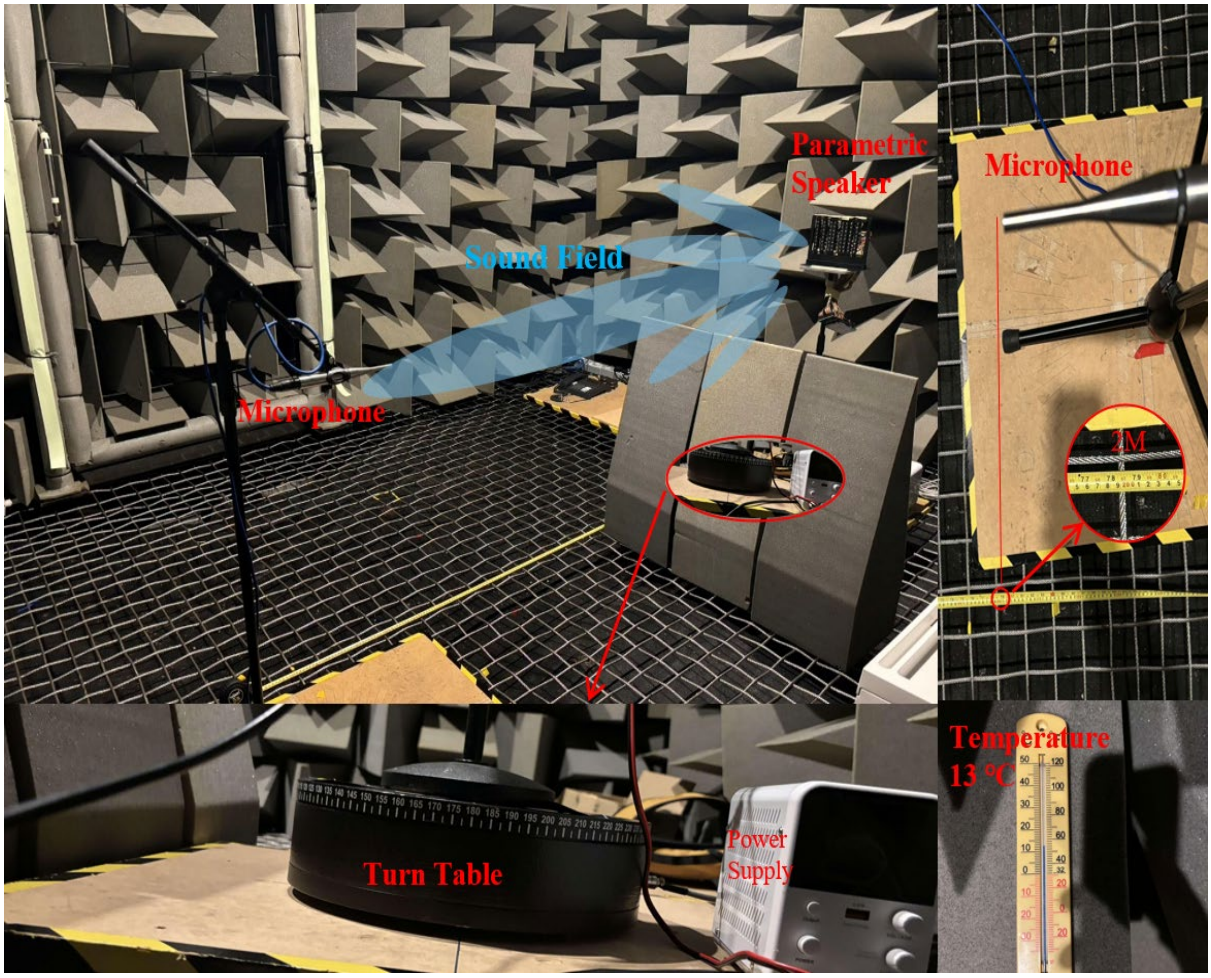


Fig. 4-7. Photograph of the experimental setup inside the anechoic chamber. The parametric array loudspeaker is mounted on a rotary stage and the microphone is fixed at a 2 m radial distance.

The PAL is mounted vertically on a rotating base positioned at the acoustic centre of the chamber, with a measurement microphone placed on a boom stand at a fixed radial distance of 200 cm along the array's main axis. Angular scanning is achieved using an electric turntable, enabling the array to rotate in 2.5° increments across a 180° arc. This configuration ensures systematic acquisition of directional sound pressure data with consistent alignment throughout each test session. Insets show the turntable, alignment of the microphone, power supply unit, and a thermometer indicating an ambient temperature of 13 °C during testing. The exact measurement geometry and environmental control parameters are summarized in Table 4-2.

Table 4-2. Detailed Parameters of Experimental Setup for Measuring Speaker Array Directivity.

Parameter	Value
Distance between speaker and microphone	200cm
Temperature T	13℃
Piezoelectric Transducer Diameter	1cm
Frequency f	40 kHz, 42 kHz
Optimal angle ψ	60°, 65°, 70°, 75°

The measurement microphone is an Earthworks M30, chosen for its linear response up to 30 kHz (with usable extension above 40 kHz), ~20 dBA self-noise, and sensitivity of −29 dB V/Pa. The microphone is mounted away from reflective surfaces and suspended above the mesh floor to minimise reflections. Signals are pre-amplified and recorded with a 24-bit audio interface and post-processed in MATLAB, following the same measurement workflow as in Chapter. 3.3.

The core acoustic source consists of an 8×8 matrix of piezoelectric ultrasonic transducers, each with a diameter of 10 mm and a nominal resonant frequency centred at 40 kHz. In the uniform configuration, the array adopts a fixed 10 mm centre-to-centre spacing between adjacent elements, corresponding to the transducer diameter. For sparse configurations, the element positions are reconfigured according to optimized spatial coordinates derived via PSO, with the objective of suppressing grating lobes and enhancing beam directivity at predefined steering angles. As shown in Fig. 4-8, a uniform linear array is used for baseline comparisons. For each optimized sparse array layout, a corresponding physical holder was fabricated using high-resolution 3D printing to ensure precise element placement. The piezoelectric transducers were manually inserted into the printed frame according to the computed coordinates. This fabrication approach enables consistent implementation of complex array geometries, facilitating accurate experimental validation of simulation and optimisation results.

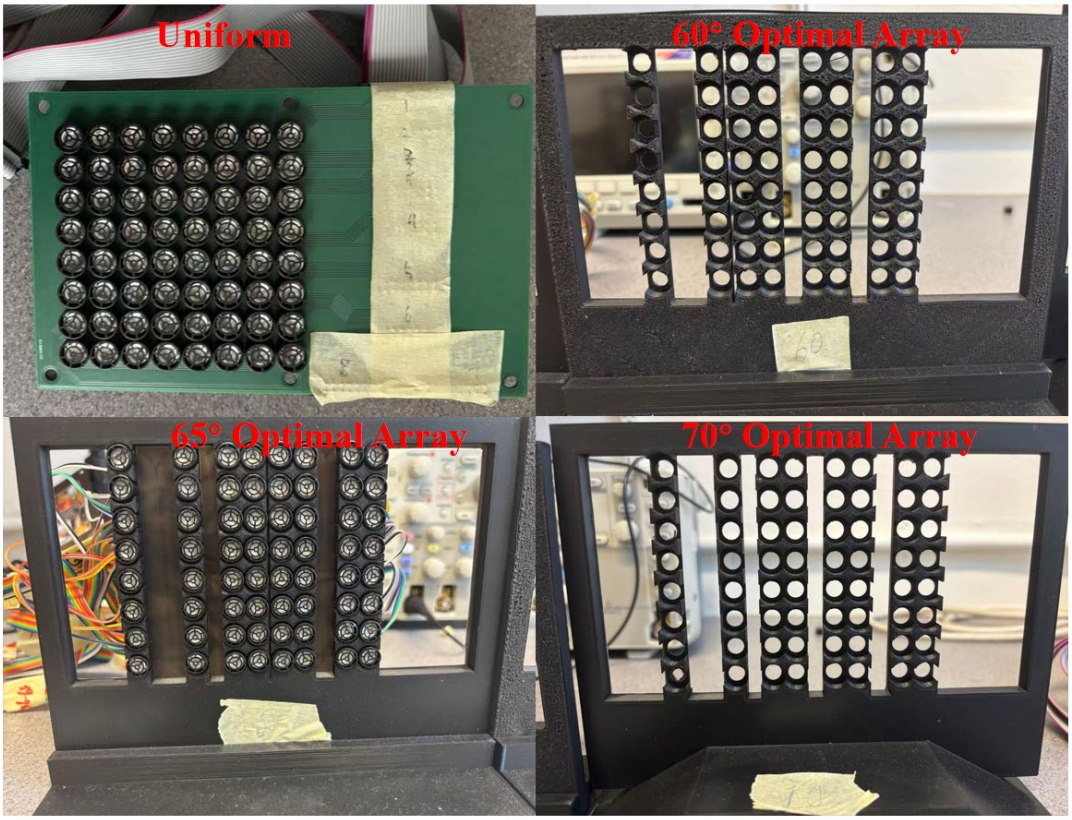


Fig. 4-8. Photograph of the physical transducer array configurations. 8 × 8 Uniform array with fixed 10 mm inter-element spacing and three optimal arrays at specific steering angles.

4.3.3 Signal Generation and Beamforming Control

The phased array loudspeaker system is driven by an Arduino Mega microcontroller, capable of generating independently controlled digital signals on any of its 64 output pins. This feature allows for precise per-channel phase modulation across the array. In the present experiment, the array is divided into eight vertically arranged groups, with each group consisting of eight transducers driven in phase synchrony. Thus, only eight independent control signals are required, one per group, which simplifies the system architecture without compromising beamforming capabilities.

Each control signal is generated as a half- duty square wave, with phase shifts implemented in 20 discrete steps ($\pi/10$ resolution) via cyclic binary pattern shifting, and amplitude controlled through duty cycle modulation. The square wave signals are amplified to 10 V and drive the piezoelectric transducers near their resonance, resulting in acoustic sinusoidal output despite the binary excitation.

The modulation scheme and signal generation process are illustrated in Fig. 4-9, which depicts the mapping between phase and duty cycle settings, controller output waveforms, and the corresponding measured acoustic responses. This system configuration enables high-resolution experimental evaluation of the directivity characteristics of both uniform and optimal PAL arrays under controlled acoustic conditions. The platform ensures consistent signal delivery, environmental repeatability, and full compatibility with simulation validation and optimisation assessment workflows.

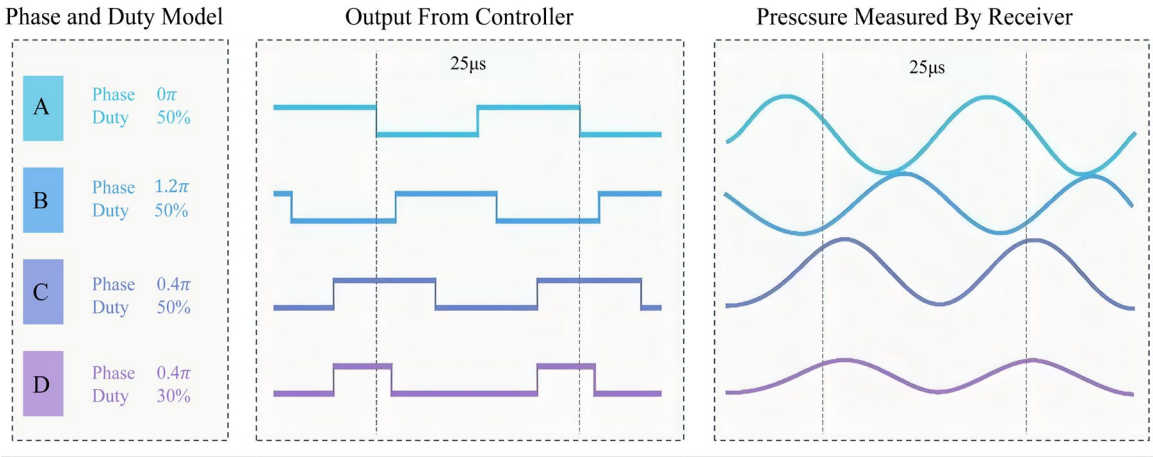


Fig. 4-9. Phase and duty-cycle control of excitation signals. Left: digital bitstreams with varying phase and duty settings. Centre: amplified signals from controller. Right: pressure waveforms captured by receivers, demonstrating sinusoidal output due to transducer resonance.

4.4 Simulation and Measurement Results for Optimized Array

This section presents the results of both simulation and experimental evaluations of transducer array configurations optimized using the PSO algorithm. The analysis focuses on comparative assessments of SPL distributions, far-field directivity patterns, and the influence of array geometry on beamforming performance. Special attention is given to the improvements achieved through PSO-based sparse arrays, including enhanced main lobe focus and side lobe suppression. These findings provide quantitative insights into the effectiveness of optimisation strategies and their implications for the design of high-performance phased array loudspeaker systems.

4.4.1 FEM Simulation Result:

Building upon the simulation and measurement setup described earlier, this section presents the results of FEM-based optimisation for different steering angles. The analysis focused on evaluating optimal array configurations and assessing directivity performance across a range of steering angles. To avoid convergence to suboptimal local minima, a random perturbation mechanism was introduced into the PSO process. Particles that failed to improve over multiple generations were perturbed to enhance exploration. This strategy led to consistent convergence behaviour across repeated runs, validating the stability and robustness of the optimisation method [120].

This optimisation process involves adjusting the inter-element spacings denoted as $d_1, d_2, \ldots, d_7$, which represent the relative distances between adjacent transducers in the sparse array. These spacings are adaptively tuned to achieve the desired directivity at each target angle ψ while minimizing side lobe levels. However, the deflection capability of a sparse array is inherently constrained by its geometry. Beyond a certain steering angle, further steering may result in reduced performance, particularly in side lobe suppression and beam shaping accuracy.

The directivity of the optimized array was initially assessed at optimal angles $\psi = 80°, 75°, 70°, 65°, 60°$ and compared with a sparse array. In the context of these arrays, 90° denotes the reference angle (zero steering). Using the optimisation algorithm, we identified the optimal transducer positions for distinct optimal angles, the iteration process is illustrated in Fig. 4-10. Convergence to the optimal solution is shown by darker solid lines, while lighter dashed lines correspond to suboptimal (local-optimal) solutions arising during iteration.

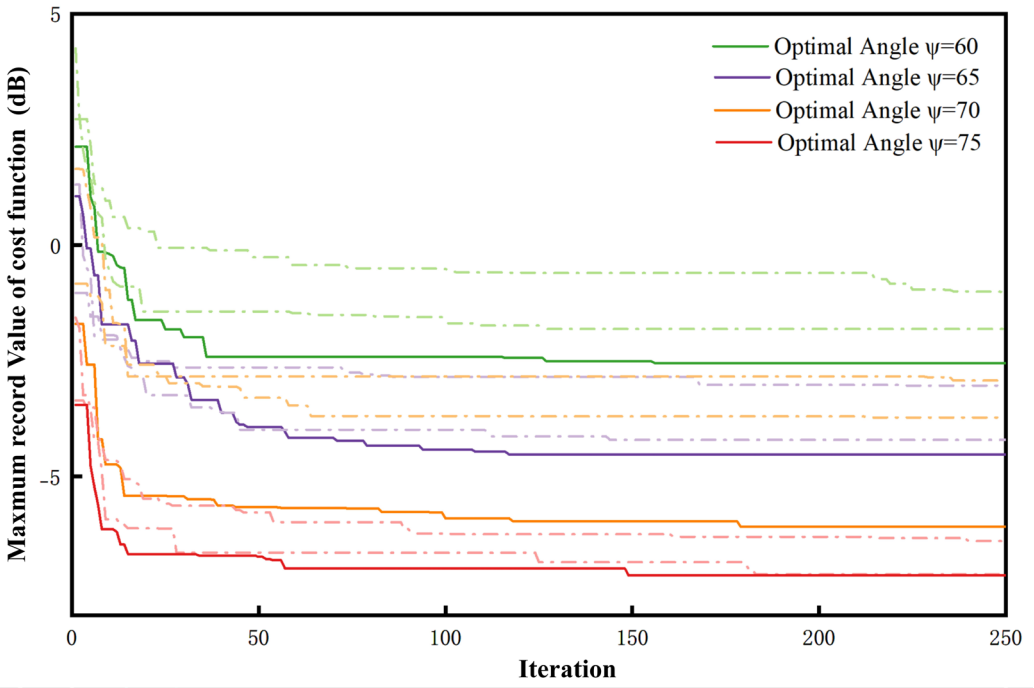


Fig. 4-10. Iterative process of Optimized PAL as a function of the number of PSO iterations, based on 250 trials. The configuration includes 8 elements ($N = 8$). Results are shown for optimal angles of: ψ = 75°, 70°, 65°, 60°.

In Fig. 4-10, each colour scheme represents the optimisation process corresponding to a specific angle. The darker data points indicate the optimal solutions identified during the process. The objective function values throughout the optimisation are negative, and the process aims to minimize these values. It is observed that as the steering angle increases, both the initial cost function value and the maximum optimized value also increase. Additionally, the optimisation process tends to converge toward the maximum value within approximately 100 iterations, suggesting that 250 iterations are sufficient to achieve an optimal solution. Furthermore, optimisation processes associated with smaller steering angles demonstrate a more favourable optimisation landscape compared to those with larger steering angles. Fig. 4-10 demonstrates that due to the random initialization of arrays at the start of each optimisation, certain local optimisation results vary during the process. This indicates a tendency for the algorithm to approach a global optimum, even when the process begins with different initial conditions. However, a general trend is evident: for each optimisation angle, the solutions consistently converge toward an optimal value.

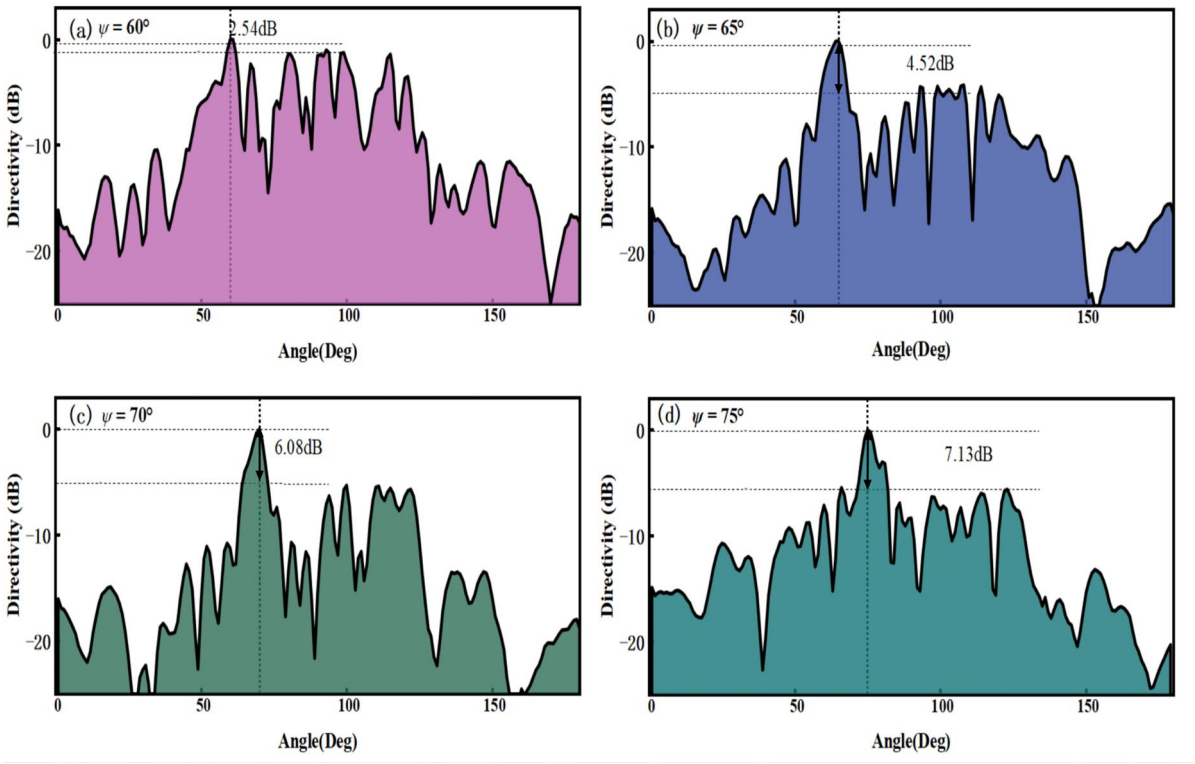


Fig. 4-11. Far-field directivity (SPL, dB) of the PSO-optimized 8-element linear array (element width =1 cm) at 40 kHz. Panels (a)-(d) correspond to target/optimal angles $\psi=\theta$=60˚,65˚,70˚,75˚ , respectively.

The patterns for optimal angles $\psi=\theta$=60˚,65˚,70˚,75˚ are shown in Fig. 4-11. The colour variations represent the results of independent simulations conducted under the same conditions. Panels (a)-(d) correspond to optimal angles $\psi=\theta$=60˚,65˚,70˚,75˚ , respectively. Curves are shown as angle-resolved profiles (0˚-180) flattened from the polar representation; the vertical dotted line marks the target steering angle ψ. For each angle, three independent simulations under identical settings are overlaid (grey traces) to illustrate repeatability; minor variations reflect stochastic restarts of the optimizer and numerical noise. All SPL values use P_{ref} =20μ Pa; when discussed comparatively, we report relative directivity $D_{\text{SPL}}(\varphi)=L(\varphi)-L(\psi)$.

As illustrated in Fig. 4-11, reveal a clear trend: as the ψ increases, the separation between the main lobe and the side lobes progressively decreases. When the ψ is below 65°, the amplitude difference between the main lobe and the side lobes becomes less pronounced, remaining under 3 dB. In the ψ range of 65° to 75°, the optimal array demonstrates effective main lobe control and side lobe suppression.

This observation suggests that the optimal array can significantly mitigate the occurrence of grating lobes. However, the deflection capability of a sparse array is inherently limited, and beyond a certain point, further deflection may result in reduced performance in terms of side lobe suppression.

Fig. 4-12 illustrates how the optimal values of (d_1,d_2,…,d_7) evolve with different steering angles, providing insight into how the array configuration dynamically adapts to directivity requirements under the limitations imposed by the sparse layout.

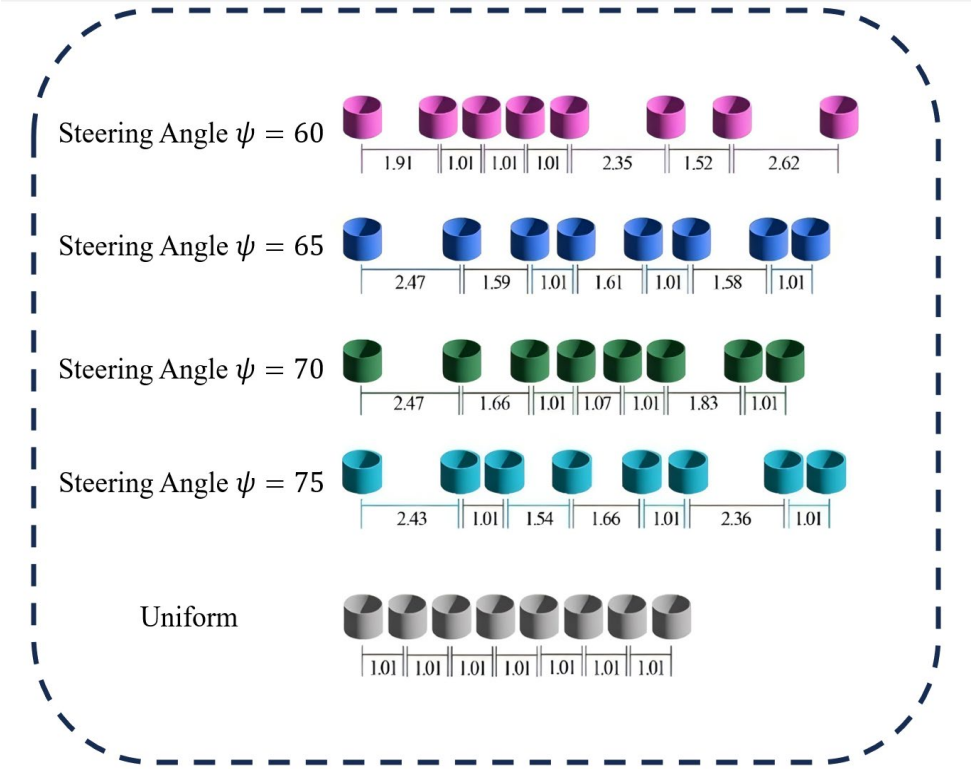


Fig. 4-12. The best array configurations obtained through optimisation for different optimal angle ψ =60˚,65˚,70˚,75˚. The array consists of 8 elements, each with a base length of 10 mm.

Table 4-3 reports the main-lobe to peak-side lobe difference (in dB) at four steering angles (ψ = 60°, 65°, 70°, 75°), comparing a PSO-optimised non-uniform array with a uniform array. Positive values mean the main lobe is higher than the strongest side lobe (better beam quality); negative values mean a side lobe/grating lobe exceeds the main lobe. Across ψ between 60° and 75°, the optimised layouts improve this metric by about +1.9 to +5.7 dB (largest gains near 60° to 70°; smaller but still positive at 75°), indicating more robust side lobe suppression under steering.

Table 4-3. Main-lobe to peak-side lobe difference (dB) for optimised and uniform arrays at steering angle ψ.

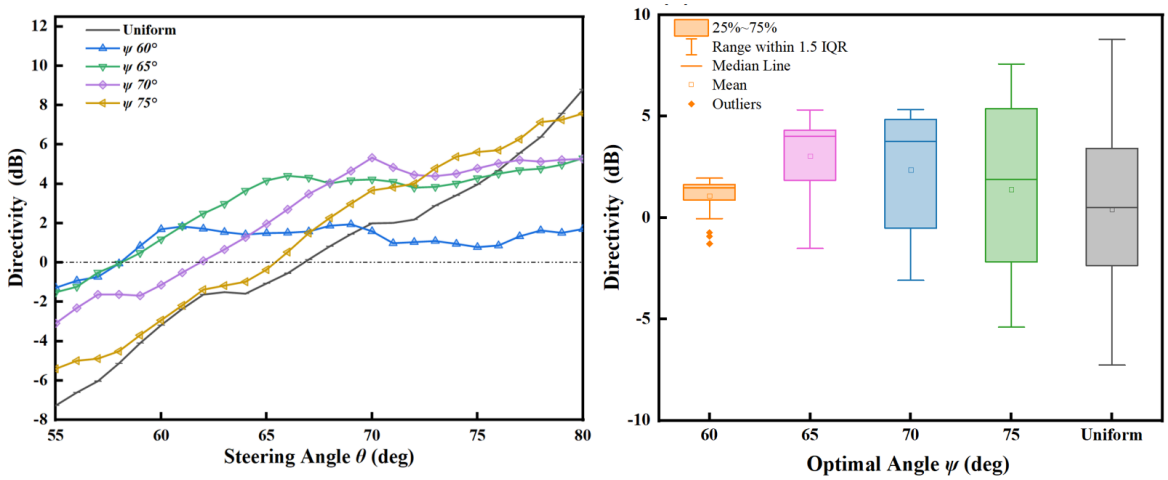
ψ (°)	Optimal array(dB)	Uniform array (dB)
60	2.5486	-3.1203
65	4.5273	-0.7920
70	6.0791	1.2819
75	7.1327	5.2665

Fig. 4-13 (a) compares the main lobe and side lobe differences (in dB) across array configurations: one Uniform and six Optimized, evaluated over steering angles (θ) from 55° to 80°. The Uniform Array maintains stable but low directivity across angles, indicating limited side lobe suppression and energy focusing. In contrast, Optimized arrays, tailored for specific angles ψ, show greater variation and improved performance at higher θ values.

For lower steering angles (θ = 55°–65°), optimized arrays show reduced distinction between main and side lobes, weakening directivity. However, performance significantly improves at higher angles (θ = 70°–75°), with stronger main lobe concentration and better side lobe suppression, especially for ψ = 65°, 70°, and 75°. This highlights the benefit of nonlinear optimisation in directional sound control. Notably, grating lobes in uniform arrays persist or even exceed side lobes at high angles, while optimized arrays effectively eliminate them, particularly around θ = 70°.

Fig. 4-13 (b) presents box plots of main lobe–side lobe differences. The Uniform Array shows a narrow interquartile range (IQR) and low median, confirming consistent but subpar performance. In contrast, Optimized arrays exhibit higher medians and wider IQRs, especially for ψ = 70° and 80°, indicating superior directivity with some trade-off in consistency due to complex non-linear geometries. The presence of outliers reflects variability tied to angle-specific optimisation.

Finally, as parametric speakers produce audible signals through nonlinear interactions of high-frequency carriers, accurate directivity modelling using the Westervelt framework requires frequency-specific analysis that accounts for both spectral characteristics and spatial propagation effects.



(a)……..……(b)

Fig. 4-13. (a) Comparison of main lobe and side lobe differences (in dB) for distinct array configurations, including a Uniform Array and six Optimized arrays, steering angle θ =60˚,65˚,70˚,75˚. (b) Comprehensive summary of main lobe and side lobe differences for each configuration across all steering angles, with each box depicting the IQR of the measured differences.

4.4.2 Experimental Validation of Directivity:

a) Measured Directivity of PALs

To experimentally validate the simulation results and assess the effectiveness of the array optimisation strategy, directivity measurements were performed for both uniform and optimized array configurations. The uniform array serves as a baseline configuration, with evenly spaced transducers, while the optimized array incorporates the sparse transducer positions (d_1,d_2,…,d_7) derived from the PSO-based simulations in Fig. 4-12.

Fig. 4-14 reports measured far-field SPL patterns at two carrier frequencies (f=40 kHz and f=42 kHz) for steering angles ψ={60˚,65˚,70˚,75˚}. Panels are mapped as follows: (a) uniform, f=40 kHz; (b) optimized, f=40 kHz; (c) uniform, f=42 kHz; (d) optimized, f=42 kHz. Curves are sampled every 2.5˚ (motorized turntable); the vertical dashed line marks ψ. The optimized layouts implement the sparse spacings (d_1,…,d_7) from Fig. 4-13. For Example, at ψ=70˚ and f=40 kHz (compare panels (a) vs. (b)), the main-peak side lobe difference improves from 1.2819dB (uniform) to 6.0791dB (optimized), evidencing tighter main lobe focus and side lobe suppression (Table 4-3).

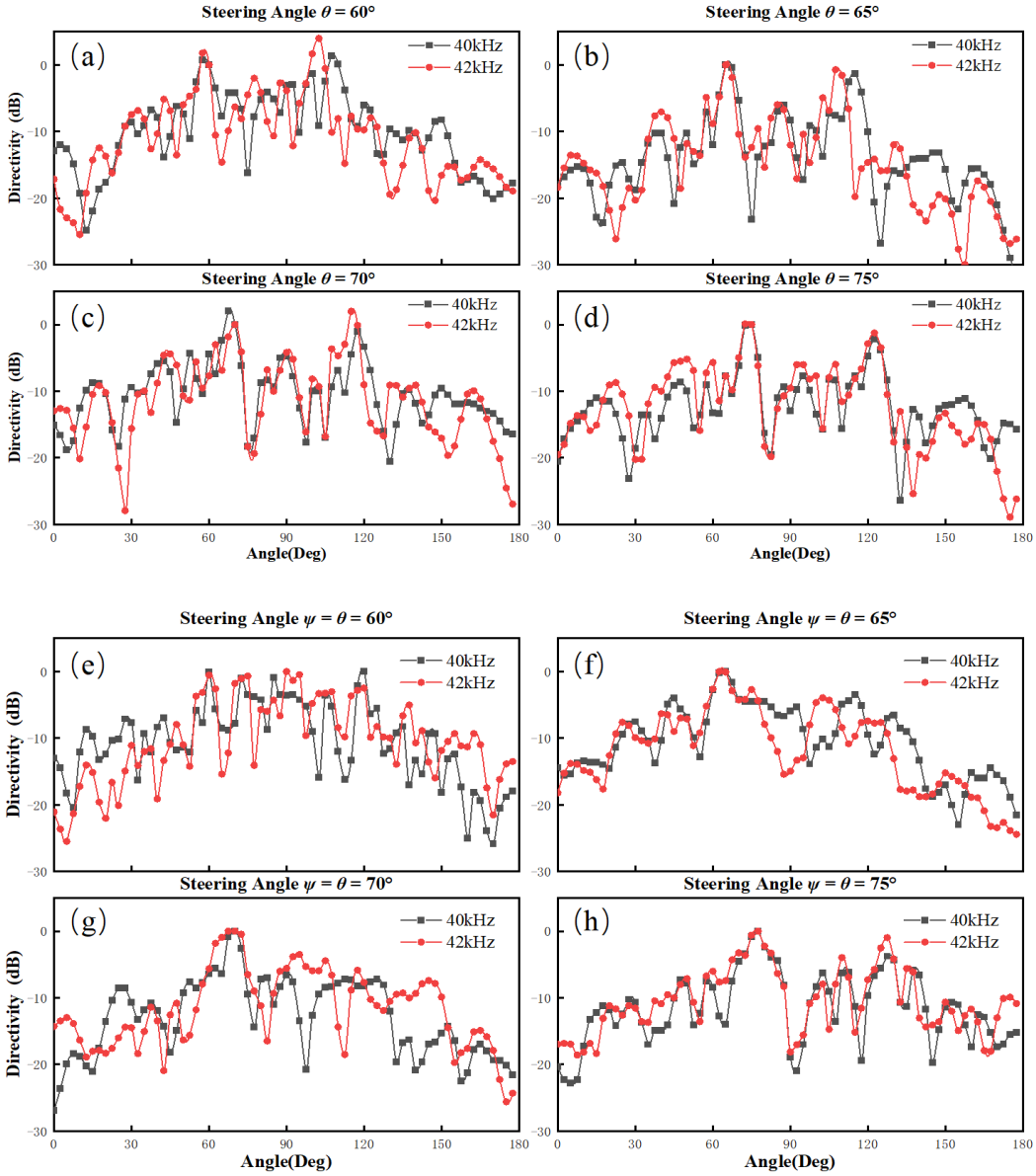


Fig. 4-14. (a-d) Directivity of the uniform array evaluated at 40 kHz and 42 kHz for steering angles ψ=60˚,65˚,70˚,75˚. (e-h) Directivity of the optimized arrays at the same frequencies and steering angles.

Experimental measurements were conducted using a PWM-driven excitation signal (50% duty cycle) powered by a 14 V supply. Directivity data were collected at an angular resolution of 2.5°, limited by the mechanical increment of the motorized turntable. Arrays corresponding to steering angles of 60°, 65°, 70°, and 75° were selected for detailed experimental analysis and validation. The setup, including equipment layout and testing environment, is thoroughly described in Chapter 4.3. The 14 V supply ensured consistent array operation, while the 2.5° sampling resolution enabled accurate capture of the directivity patterns.

The experimental results show the directivity performance at f= 42 kHz, 40 kHz for the uniform array, the optimal ψ = 60° array, and the optimal 75 array. Compared to the simulation results, the experimental side lobes are higher, and some irregular side lobes appear. These discrepancies are likely due to performance deviations among the transducers, measurement errors in the array setup, and phase errors in the control circuitry. However, the experimental results still validate the key conclusions drawn from the simulations.

The experiments show that nonlinear optimisation can suppress sidelobe levels at selected steering angles. At 75°, the optimised array produces visibly smaller grating lobes than the uniform array. Its pattern is also smoother, which matches the simulation trends. At high steering angles, the optimised array keeps high main-lobe directivity, though sidelobes unavoidably rise beyond 65°.

b) Westervelt Directivity of PALs

A detailed analysis of Fig. 4-15 visualises the Westervelt directivity for multiple array layouts at steering angles of 75°, 70°, 65°, and 60°. In each subplot, the experiment results(blue) are plotted against simulation results(red) for a direct comparison. Both the simulation and the experiment use the primary field only to compute Westervelt directivity via the $D_1 \times D_2$ formulation (i.e., derived from the first-wave response). The results show that the optimised arrays form sharper main lobes and achieve stronger sidelobe suppression, with a clear advantage at 60° steering. This confirms the effectiveness of the optimisation strategy.

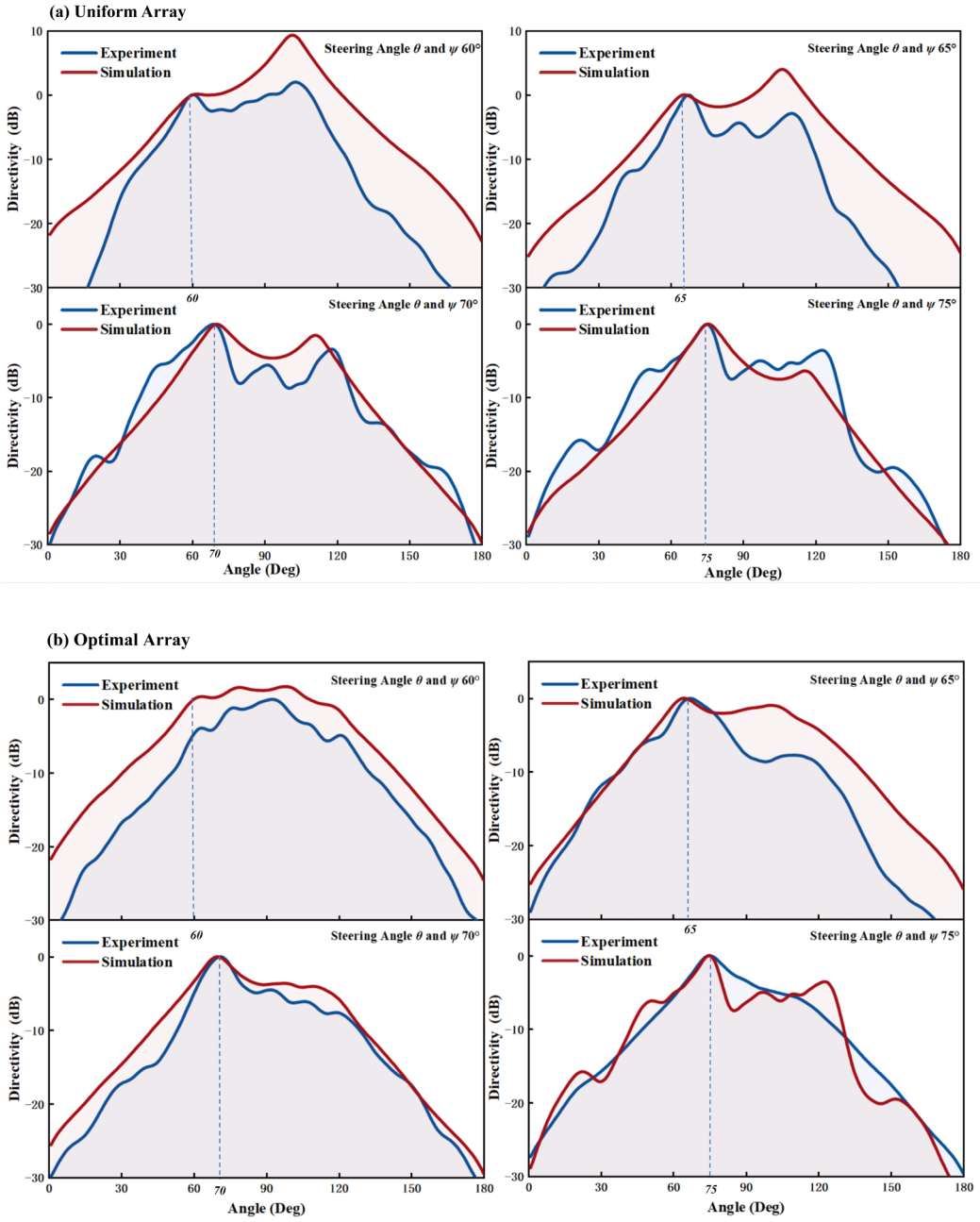


Fig. 4-15. FEM Simulated and measured Westervelt directivity of the uniform array(a) and optimal array(b) at 40 kHz and 42 kHz for steering angles ψ=60˚,65˚,70˚,75˚;

By analysing the Westervelt directivity, it becomes evident that the optimal array produces smoother patterns compared to the uniform array. At steering angles of ψ = 65° and 70°, the optimal array clearly outperforms the uniform arrays by effectively reducing side lobes while maintaining strong main lobe directivity. At ψ = 75°, the performance of the optimal and uniform arrays is comparable. In contrast, at ψ = 60°, the optimal array loses its advantage in main lobe strength, and both uniform arrays demonstrate higher directivity, particularly at the

grating lobe angles. These observations suggest that the optimal array offers significant benefits in the mid-to-high steering range (ψ = 65° to 75°), where it suppresses grating lobes while preserving the main beam shape. However, its effectiveness diminishes at lower steering angles, where grating lobes are less prominent and the performance gap between configurations is minimal.

The consistency between experimental and simulation results highlights the reliability of the optimisation method in reducing unwanted grating lobe interference, thus improving overall system performance at these target angles. However, from the perspective of practical application, only arrays with lower optimal steering angles are meaningful. When the optimal steering angle is set too large (e.g., 60°), the array cannot maintain high main lobe directivity at low steering angles, nor can it significantly reduce grating lobes at high steering angles, resulting in a lack of distinctiveness in the main lobe.

However, some discrepancies between simulation and experimental data are observed. The experimental results tend to show higher side lobes and irregular patterns, which are attributed to transducer inconsistencies, phase control errors, or minor misalignments in the physical array setup. These deviations are expected in practical implementations and underscore the need for further refinement in hardware synchronization and experimental calibration. Nonetheless, the overall trend remains consistent, validating the simulation models and confirming that the optimisation process is effective in guiding real-world array design.

Table 4-4 presents a comparative analysis of the directivity differences between the main lobe and grating lobes for both optimized and uniform arrays at different steering angles (ψ = 60°, 65°, 70°, and 75°). The table includes both simulation and experimental results, demonstrating the performance improvements achieved through optimisation. The optimized arrays consistently show higher main lobe-to-grating lobe differences, particularly at larger steering angles, confirming their superior directivity control. The comparison between simulated and experimental data further validates the effectiveness of the proposed optimisation approach in real-world implementations.

Table 4-4. The directivity difference between main-lobe and grating lobe at optimal steering angle ψ and comparison with the results of linear array.

	Simulation Data (dB)		Experiment Data (dB)	
ψ (°)	Uniform Array	Optimal Array	Uniform Array	Optimal Array
60	-9.2042	-1.539	-1.966	-3.8191
65	-4.1416	0.917	2.950	5.88176
70	1.4966	3.6541	3.449	4.53129
75	6.46171	5.6765	3.538	4.47579

While the consistency between simulation and experimental results supports the effectiveness of the proposed optimisation approach, several sources of measurement error and modelling assumptions must be acknowledged, as they may contribute to the observed discrepancies.

On the experimental side, hardware-related limitations include variations in microphone sensitivity, imperfections in transducer responses, and possible phase mismatches in signal excitation. Minor misalignments during array installation, such as deviations in element orientation or positioning, can introduce systematic errors in the measured directivity. In addition, the angular resolution of the measurement system is limited to 2.5°, which restricts the ability to capture fine-scale variations in the side lobe structure, especially near nulls or secondary peaks.

Environmental factors also play a role. Despite efforts to minimize reflections using absorbing materials, residual acoustic reflections and background noise may affect measurement reliability, particularly at higher frequencies or lower signal amplitudes. Furthermore, variations in temperature and humidity can alter the local speed of sound and air absorption characteristics, which are not dynamically compensated for in the current setup.

On the simulation side, several idealizations are made to maintain computational efficiency. These include the assumption of homogeneous and lossless air, ideal phase synchronization across array elements, and the use of simplified 2D models for certain cases. While these approximations are acceptable for initial design and comparative analysis, they may overlook important 3D propagation effects and hardware-specific nonlinearities.

Addressing these issues through more precise alignment mechanisms, higher-resolution angular scanning, dynamic environmental monitoring, and enhanced physical modelling would further improve the accuracy and robustness of future directivity evaluations.

4.5 Summary

This chapter evaluated array optimisation for PAL directivity using a FEM–PSO–experiment pipeline. Both simulation and measurement computed Westervelt directivity from the primary field via the Westervelt formulation. Tests covered steering angles ψ**=60˚,65˚,70˚,75˚**at 40/42 kHz, with 2.5° angular sampling. Across these settings, optimised sparse arrays produced sharper main lobes and lower sidelobes than uniform arrays. For example, at ψ**=70˚**, 40 kHz, the main-peak–sidelobe difference improved from 1.28 dB (uniform) to 6.08 dB (optimised). Table 4-4 and Fig. 4-15 show the same trend: optimised layouts raise the main-to-grating-lobe gap, especially at 65°–75°.

Experiments and FEM agreed on the steered main-lobe angle, grating-lobe locations, and the relative benefit of optimisation. Measured sidelobes were higher and sometimes irregular, attributable to element mismatch, small phase errors, and minor alignment drift. These deviations are expected in practice and point to the value of tighter hardware synchronisation and calibration. Still, the agreement in trends confirms that the optimisation reliably reduces grating-lobe energy and improves beam focus at the target angles.

Practically, optimisation benefits depend on the intended steering range. Mid-to-high steering (65°–75°) showed the clearest gains; at 60° the advantage in main-lobe strength can diminish, and uniform arrays may retain comparable directivity at some angles. Thus, the optimisation target ψ should match the operational band: targeting very high angles can trade off performance elsewhere. Overall, the FEM-guided, PSO-optimised designs provide measurable sidelobe suppression and cleaner beams under the tested conditions, offering a useful route for PAL array configuration in practice.

Chapter 5

GAN-based Hybrid Learning Framework for Far-Field Directivity Prediction in PALs

FEM simulations in Chapter 4 give high-fidelity PAL directivity. They require fine spatial discretisation at ultrasonic frequencies. This leads to high computational cost. They are not suitable for interactive design or real-time control. To address this, we build a data-driven surrogate. The surrogate maps near-field representations to far-field patterns.

In this chapter, we adopt a conditional Generative Adversarial Network (GAN): the generator takes the input and predicts a far-field map. The discriminator checks whether the map looks real when given the same input. We train the two networks with an adversarial loss and a reconstruction loss. These losses keep the output sharp and physically plausible, and they stabilise training. After training, prediction reduces to a single forward pass with negligible marginal cost, enabling fast design evaluation and adaptive beamforming.

5.1 Introduction

5.1.1 Motivation for GAN-Based Prediction

Accurate modelling of the parametric speaker's sound field is essential for optimising directivity and improving audio performance. Analytical methods, introduced in Chapter 3, give fast estimates. However, they do not capture key nonlinear effects and complex wave interactions that shape far-field behaviour.

FEM-based models, discussed in Chapter 4, are accurate and flexible. They handle complex geometry and boundary conditions well. However, they are costly, especially for 3D time-domain runs. The solver must discretise the mesh, perform numerical integration, and solve large sparse matrices. These steps need high-performance computing and long runtimes. They are not practical for large parameter studies or real-time use.

This limitation motivates an alternative that keeps accuracy and lowers cost. Recent deep-learning methods, especially GANs, offer such a path. GAN-based models learn the nonlinear map from near-field data to far-field directivity. They deliver rapid predictions without repeated full-physics simulations [96], [119]. In contrast to the FEM, which involves mesh discretization, numerical integration, and the solution of large sparse matrices resulting in high memory usage and exponentially increasing computation time, GAN inference is significantly more efficient. It can be executed on a single GPU within milliseconds and with much lower energy consumption [95]. This makes GAN prediction suitable for real-time tasks such as adaptive beamforming and dynamic array control. Table 5-1 compares the computational cost and resource needs of FEM and GAN-based methods.

Table 5-1. Comparative Analysis of Computational Methods for Acoustic Field Prediction: FEM Simulation vs. GAN-Based Prediction.

	Computing Complexity	Computing Resource	Error
FEM Simulation	$O(N^P)(p=2\text{-}3)$	High	High quantifies
GAN Prediction	$O(M)$ (fixed)	Low	Potential generalization error

This study develops a GAN-based framework to predict the far-field directivity of parametric array loudspeakers (PALs) with high accuracy while reducing reliance on computationally expensive full-wave simulations. Finite-element modelling (FEM) is used to generate high-fidelity supervisory data and to benchmark behaviour under nonlinear propagation and realistic geometries. By coupling data-driven learning with physics-based simulation, the approach enables rapid, measurement-referenced evaluation and streamlines PAL design and optimisation for practical deployments.

5.1.2 Selection of GAN Architectures for Spatial Sound Field Prediction

This study predicts far-field acoustic distributions at 200 cm from near-field data measured at 30 cm. The task is high-dimensional. Propagation and wave interference make the mapping complex. Traditional regression lacks the nonlinear capacity needed for these interactions. Standard deep networks often over-smooth spatial detail. These smoothing lowers resolution and degrades the predicted field.

We adopt a GAN framework to learn far-field patterns with preserved spatial detail and improved data efficiency. High-fidelity FEM simulations provide supervisory signals and a physics-consistent reference for benchmarking nonlinear propagation and realistic geometries. Compared with running FEM at design time, a trained GAN generates patterns at interactive rates, enabling rapid, measurement-referenced evaluation within design workflows[121], [122].

GAN is a training framework, not a fixed network. The generator can use different designs. In this study, we test two generators: U-Net and an Encoder–Decoder. We compare their ability to predict far-field acoustic directivity.

U-Net uses an encoder–decoder with skip connections. It is common in image-to-image translation. It preserves spatial detail and global context. This makes it a strong choice for near-field to far-field mapping. We also train a supervised U-Net as a baseline to measure the gain from GAN training [123]. The Encoder–Decoder compresses the input to a latent code and reconstructs the output. It offers flexible feature extraction and can generalise better with limited data. Used as a GAN generator, it helps reduce overfitting while maintaining accuracy [124].

Additionally, We also consider CNNs and Transformers. They offer different trade-offs between computational cost and feature representation. These architectures are included in Table 5-2, which presents a comparative analysis of different deep learning models for acoustic field prediction.

By integrating GAN with U-Net and Encoder-Decoder and comparing them to a standalone U-Net, this study assesses the impact of different generator designs on accuracy, efficiency, and generalization in nonlinear acoustic field prediction. The comparative analysis in Table 5-2 further contextualizes these findings by evaluating additional architecture, providing a comprehensive overview of deep learning approaches in this domain. The insights gained contribute to the development of a GAN-based framework for parametric speaker modelling, advancing deep learning-based sound field prediction and optimisation. The following section details the methodology used to implement and evaluate these models.

Table 5-2. Comparative analysis of deep-learning architectures for acoustic field prediction.

Model	Reference	Advantages	Limitations
U-Net	[123], [125]	Strong spatial detail via skip connections; good multi-scale dense mapping.	Limited global context; may smooth high-frequency content.
Encoder-Decoder	[124]	Simple, efficient, scalable; good for resolution change.	Loses fine detail without skips; deconvolution artifacts possible.
CNN (generic/FCN)	[126]	Fast inference; stable training; local pattern capture.	Weak long-range dependencies; needs dilations/multi-scale.
Transformers	[127], [128]	Global context; strong transfer/self-supervised pretraining.	Data-/compute-hungry; weaker inductive bias in small-data regimes.

5.2 GAN-Based Methodology for Far-Field Directivity Prediction

In this chapter, we collect acoustic-field data in a defined region and split it into near-field and far-field sets. We use a 2D framework with the Pressure Acoustics module to simulate sound propagation in space. The sampling domain has 512 × 512 points. We use the near-field data as the GAN input. We use the paired far-field data as the target output. This setup trains the model to learn the nonlinear propagation behaviour. We set the near/far split from the physics of the difference-frequency (DF) wave. We estimate the onset of the far field as follows:

$$r_{\text{far-field}} = \frac{2D^2}{\lambda} = \frac{2\times(0.15)^2}{0.171} \approx 0.117\ m \tag{5-1}$$

Where $\lambda = c/f$, c is the speed of sound in air (approximately 342 m/s), and f is the frequency of the difference-frequency wave. Assuming the two primary high-frequency ultrasound waves have frequencies of 40 kHz and 42 kHz, resulting in a difference-frequency wave of 2 kHz. The effective diameter D of the array is estimated to be approximately 15 cm. This is based on a linear arrangement of eight transducers, each with a diameter of 1 cm, spaced with small gaps in between. Although the exact centre-to-centre spacing may vary slightly due to mounting constraints, the total aperture spans a relatively wide area, forming a compact but acoustically

significant array. To avoid losing near-field information, the near field is defined up to 30 cm, and the transition to the far-field is assumed to begin beyond 200 m.

Fig. 5-1 compares FEM-based far-field simulation with the proposed GAN-based estimation. The left panel shows the FEM model. It resolves the primary wave, the secondary wave, and the grating lobes across a 200 cm semi-circular domain. The right panel shows the GAN approach. It takes near-field data on the 30 cm semi-circular arc as input and predicts the far-field directivity.

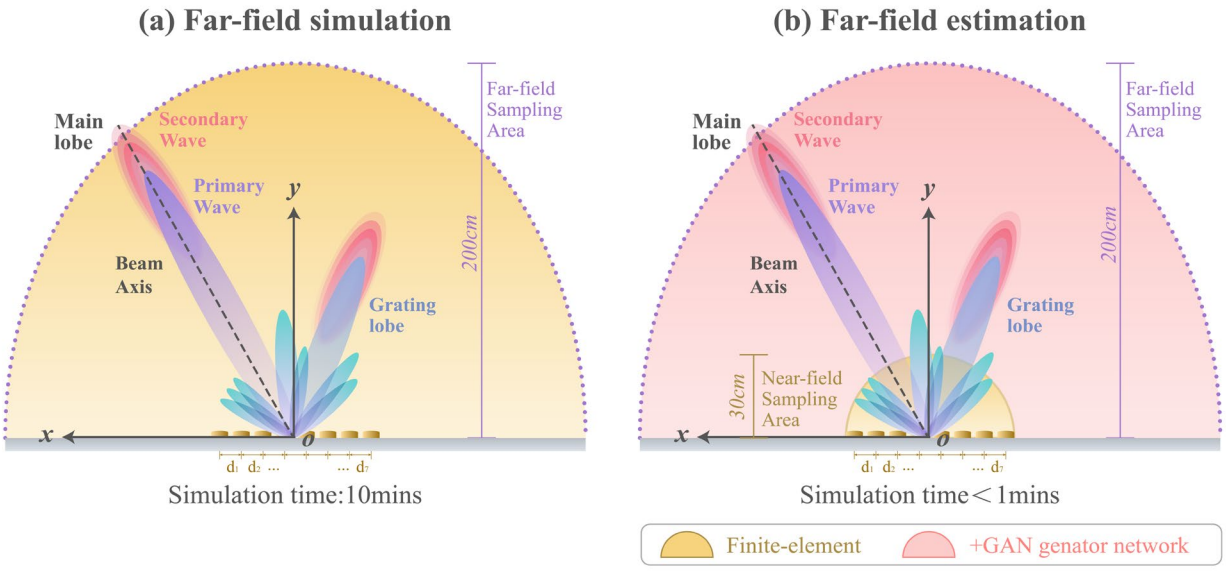


Fig. 5-1. Far-field estimation using a GAN compared to finite-element simulation, (a) Finite-element simulation method, (b) Near-field simulation with far-field estimation.

5.2.1 Data Collection and Preprocessing

In this simulation, the directivity patterns of an ultrasonic phased array were calculated using the Westervelt equation within a finite element framework. To reduce computational complexity, a 2D axisymmetric model was adopted to represent the inherently 3D system. This was achieved by placing a line source along the axis of symmetry using a built-in on-axis node in the 2D axisymmetric geometry.

The directivity patterns were computed using a frequency-domain finite-element formulation identical to Chapter. 4 (Equations. (4-1) - (4-6)), i.e., a Helmholtz model with Westervelt-

informed weak nonlinearity truncated at the second harmonic. We therefore do not repeat the governing equations, symbols, or unit definitions here; notation and units.

To reduce computational cost while retaining principal-plane accuracy, we adopt the same **2D** axisymmetric surrogate used previously. The phased aperture is represented by a phased line source distribution placed on the symmetry axis (COMSOL on-axis line source), which corresponds to a radially vibrating cylindrical emitter of vanishing radius. Steering is implemented by prescribing relative phase delays across eight coaxial source segments located along the axis, emulating an eight-element linear array in the principal plane. This surrogate captures the dominant beam physics for the slender apertures and steering angles considered; see Chapter. 4.2 for the 3D extension and its computational implications.

This Fig. 5-2 uses the same COMSOL Pressure Acoustics (Frequency Domain) configuration as Fig. 4-2. A 2D principal-plane semicircular region of radius 200 cm is terminated by a PML to suppress artificial reflections. Near- and far-field evaluation arcs are placed at 30 cm and 200 cm, respectively. The far-field radius matches the experimental microphone distance and lies well beyond the Rayleigh region at 40–42 kHz, making the angular pattern effectively range-invariant. The mesh policy is identical (minimum element size $\geq \lambda_{\min}/6$) to ensure accuracy and consistency across figures. All other solver, boundary, and normalization settings follow those reported for Fig. 4-2.

Post-processing follows chapter. 4.2: total acoustic pressure p_t is obtained from the finite-element solution, SPL is computed using the standard reference $P_{\text{ref}}=20\mu$ Pa, and relative directivity is reported as $D_{\text{SPL}}(\varphi)=L(\varphi)-L(\varphi_{\text{ref}})$ with φ_{ref} the main-lobe direction.

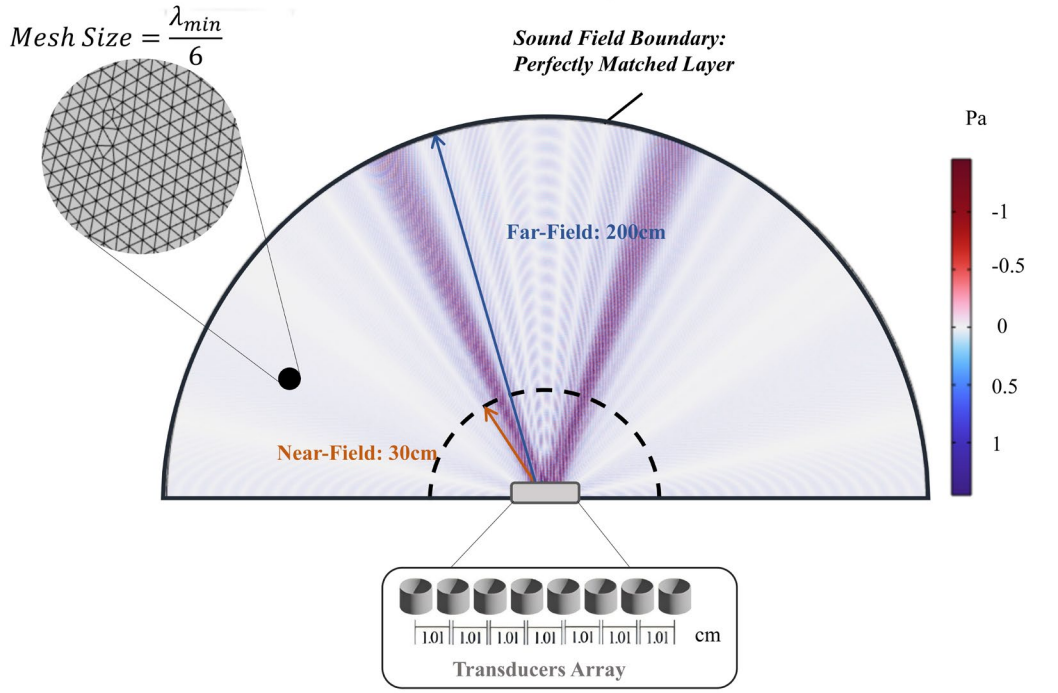


Fig. 5-2. Far-field sound pressure simulation of a parametric array loudspeaker array conducted in COMSOL.

We construct a 512 × 512 grayscale image from FEM outputs by sampling SPL over a 180° sector. Each pixel encodes the SPL at one spatial location. The grayscale intensity reflects the local pressure magnitude. This step converts the complex sound field into an image-like input. It suits deep-learning models that extract spatial features and recognise patterns.

To improve robustness and generalisation, we use a hybrid data-generation strategy. We run grid scanning to sample key array configurations at small steering angles. We add random sampling to vary transducer positions, amplitudes, and phases. We also apply standard image augmentation, such as random cropping and flipping, in line with prior image-learning results [129]. Structural perturbations were introduced based on the recommendations of Shorten and Khoshgoftaar [130], who showed that targeted data augmentation can significantly improve model performance under limited data conditions.

The dataset consisted of 5,000 samples: 4,650 simulation instances and 350 experimental measurements. In the simulations, each “array” denotes a distinct geometric realisation of a linear transducer aperture. We considered three families: uniformly spaced, Gaussian-distributed inter-element spacing, and randomly perturbed (jittered) spacing and, within each family, generated multiple arrangements by varying the element count N, aperture length A,

and spacing-profile parameters (e.g., Gaussian width). For every arrangement, far-field responses were computed over steering angles $\theta \in [60^{\circ}, 90^{\circ}]$; small random perturbations around the baseline designs were added to increase diversity. The 350 experimental samples were reserved for validation and out-of-sample generalisation across the same angle range.

To ensure consistency across training inputs, all grayscale images were normalized to the range [0,1] using min-max scaling. Specifically, each pixel intensity I was transformed according to the Equation (5-8):

$$I_{\text{norm}} = \frac{I - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}} \tag{5-2}$$

where I_{min} and I_{max} denote the minimum and maximum intensity values in the original image, respectively. This normalization facilitates stable training and accelerates convergence during learning.

5.2.2 Deep Learning Model Development

The proposed GAN architecture, illustrated in Fig. 5-3, comprises a generator and discrimination, specifically designed for far-field acoustic prediction. During adversarial training, the generator continuously refines its predictions to deceive the discriminator, ultimately converging toward an equilibrium where the discriminator can no longer distinguish between real and generated far-field patterns [131], [132]. This enables the synthesis of high-fidelity acoustic field distributions while reducing reliance on computationally expensive simulations.

To further evaluate the effectiveness of GAN-based learning, this study implements two distinct network architectures: U-Net and Encoder-Decoder. These architectures serve as the generator backbone and are compared in terms of their ability to reconstruct far-field directivity patterns. The following sections provide a detailed discussion of these two approaches, outlining their structural differences, advantages, and suitability for nonlinear acoustic field prediction.

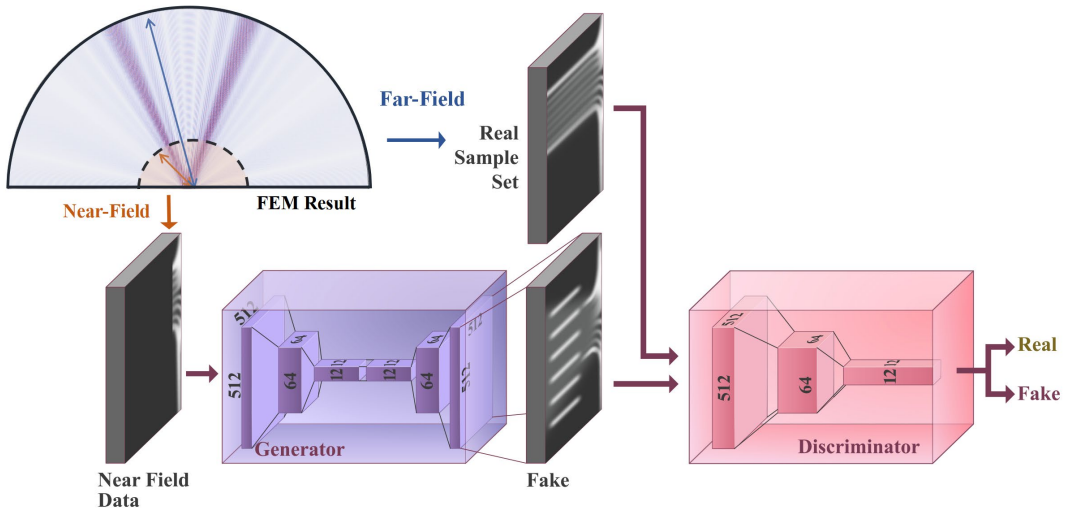


Fig. 5-3: Illustration of GAN Training Process: Interaction Between Generator and Discriminator.

a) Loss Function

To ensure the effectiveness of the GAN architecture illustrated in Fig. 5-3, a well-designed loss function is critical for achieving both training stability and predictive accuracy. The generator must produce far-field acoustic distributions that are not only visually consistent with real measurements but also physically aligned with the principles of wave propagation. To this end, the training objective integrates three complementary loss components:

- **L1 loss** (also referred to as Mean Absolute Error or MAE), which serves as the reconstruction loss by enforcing pixel-wise similarity with ground truth data.
- **Adversarial loss**, which encourages realism through feedback from the discriminator; and
- **Physics-informed loss**, which enforces consistency with the acoustic wave equation to preserve physical plausibility.

Together, these components define the generator's overall objective function, balancing perceptual fidelity with physical validity. The following sections detail the formulation and role of each loss term [133], [134].

The L1 loss measures pixel-wise similarity and ensures that generated outputs closely resemble ground truth distributions which have been defined in equation (5-3).

$$L_{\mathrm{L1}}=E_{x,y}\left[\|y_{gt}-G(x)\|_1\right]=\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i| \tag{5-3}$$

Where y_{gt} represents the ground truth far-field data, $\hat{y}_i$=G(x) is the predicted far-field output generated by the GAN, and N denotes the total number of samples.

To improve realism, an adversarial loss is used to train the generator against a discriminator. A hinge loss formulation (5-4) is adopted to stabilize training and enhance gradient quality [121]:

$$L_{adv}=E_y\left[\max\left(0,1-D(y)\right)\right]+E_x\left[\max\left(0,1+D\left(G(x)\right)\right)\right] \tag{5-4}$$

This formulation ensures that the discriminator assigns a high probability to real far-field samples while penalizing the generator when its outputs deviate from realistic distributions.

To enforce physical consistency, a physics-informed loss is introduced based on the wave equation. It ensures the generated acoustic field aligns with wave propagation behaviour [135]:

$$L_{\mathrm{acoustic}}=\mathrm{E}_x\left[\left\|\nabla^2 G(x)-\frac{1}{c^2}\frac{\partial^2 G(x)}{\partial t^2}\right\|_2\right] \tag{5-5}$$

Where c represents the speed of sound. By incorporating this constraint, the GAN-generated acoustic field aligns with the expected wave propagation behaviour, improving the physical accuracy of the far-field estimation. The overall objective function for the generator combines adversarial loss, L1 loss, and acoustic consistency loss, balancing realism and physical accuracy. The generator loss is expressed as:

$$L_{\mathrm{G}}=L_{adv}+\lambda_1 L_{L1}+\lambda_2 L_{acoustic} \tag{5-6}$$

Where λ_1,λ_2 are hyperparameters that control the relative contributions of the reconstruction and physics-based losses.

The discriminator loss L_D is designed to improve the ability of the discriminator D to distinguish between real and generated samples. The objective of the discriminator is to assign a high probability to real samples y and a low probability to generated samples $G(x)$. For a hinge loss-based discriminator, the loss function is defined as:

$$L_D=E_y\left[\max\left(0,1-D(y)\right)\right]+E_x\left[\max\left(0,1+D\left(G(x)\right)\right)\right] \quad (5\text{-}7)$$

This loss function encourages the discriminator to assign higher confidence scores to real far-field samples while penalizing incorrect classifications of generated samples. By reinforcing this distinction, the discriminator effectively guides the generator toward producing more realistic far-field distributions.

b) Encoder-Decoder:

As shown in Fig. 5-4, the generator adopts a five-stage encoder–decoder architecture with residual blocks. The encoder progressively reduces spatial resolution while increasing the number of feature maps, enabling effective hierarchical feature extraction; the decoder mirrors this with five up-sampling stages to restore resolution and fuse features via skip connections, yielding high-fidelity reconstruction. In the diagram, the numbers above each block denote the number of feature maps (channels) at that stage, and the numbers below the arrows denote the spatial resolution (e.g., 512×512). This design improves the accuracy and stability of far-field directivity predictions [123].

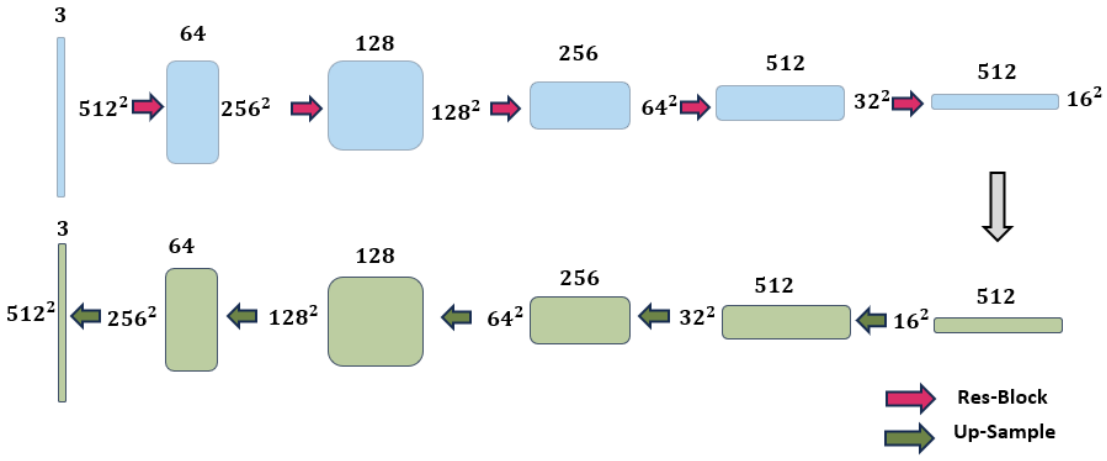


Fig. 5-4. Residual Encoder-Decoder Generator for Far-Field Acoustic Prediction.

The encoder (E) extracts hierarchical acoustic features by applying convolutional layers and progressively reducing spatial dimensions. This compression process ensures that essential information is retained while eliminating redundancies. The encoding function is formulated as Equation (5-8) [136]:

$$Z_{latent}=E(X_{input};\theta_e)=X_{input}+f\left(W_e*X_{input}+b_e\right) \quad (5\text{-}8)$$

Where X_{input} represents the input of near-field acoustic data, Z_{latent} is the latent representation, W_e,b_e are the encoder weights and biases, and $f(\cdot)$ denotes the activation function. As illustrated in Fig. 5-4, the encoder progressively reduces spatial dimensions from 512^2 to 16^2 while increasing feature depth, utilizing residual blocks (red arrows) to preserve information across layers. The decoder (D) reconstructs the far-field output from the latent representation through transposed convolutions or upsampling operations, incorporating residual connections to refine the reconstruction and maintain spatial consistency. The decoding process is expressed as:

$$\hat{Y}=D(Z_{latent};\theta_d)=Z_{latent}+g(W_d*Z_{latent}+b_d) \quad (5\text{-}9)$$

Where $\hat{Y}$ represents the reconstructed far-field output, W_d,b_d are the decoder weights and biases, and $g(\cdot)$ denotes the activation function. The decoder restores the spatial resolution from 16^2 to 512^2, using upsampling operations (green arrows) to maintain acoustic field consistency.

This residual encoder-decoder framework, when integrated into the GAN training process, enhances the generator's ability to produce realistic and high fidelity far-field acoustic predictions, aligning with the loss function objectives previously discussed.

c) Overview of U-Net Architecture

The Residual U-Net follows the Encoder-Decoder structure but integrates residual and skip connections to improve gradient flow and feature reuse which have been shown in Fig. 5-5. Unlike conventional Encoder-Decoder models, U-Net preserves fine spatial details by directly transferring encoder features to the decoder, making it well suited for high-resolution acoustic modelling.

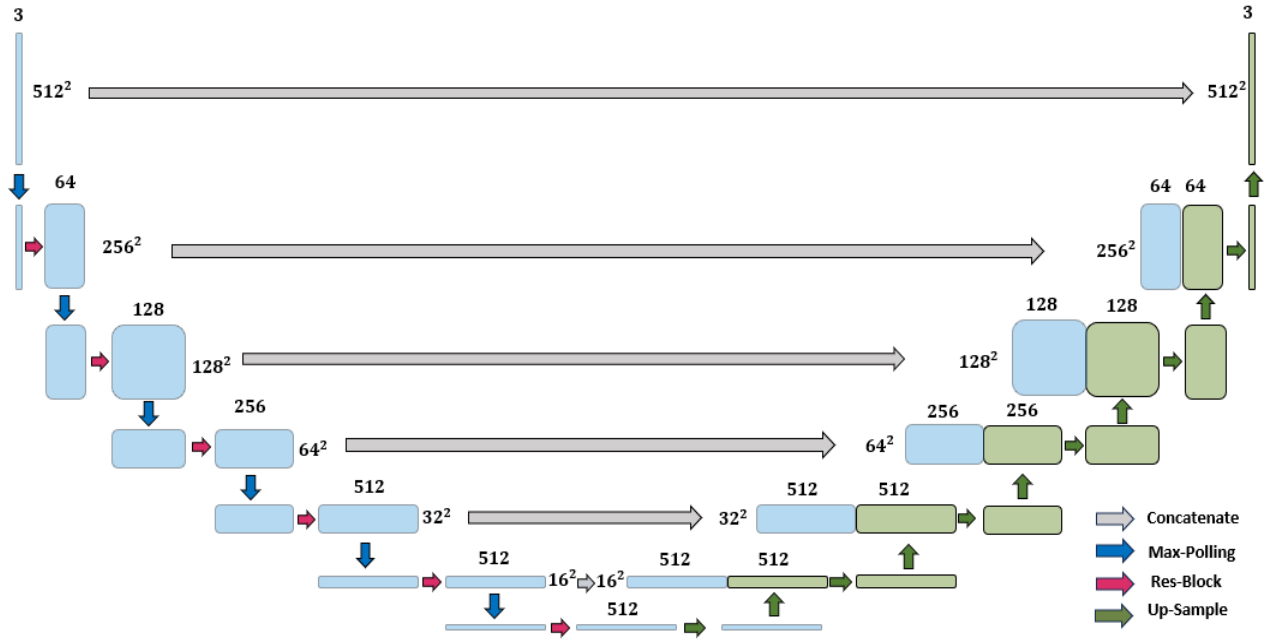


Fig. 5-5. Residual U-Net Architecture for Far-Field Acoustic Prediction.

The encoder extracts hierarchical features through residual convolutional blocks while down sampling the spatial resolution via max pooling. The encoding process is expressed as [137]:

$$Z_l = Z_{l-1} + f\left(W_{e,l} * Z_{l-1} + b_{e,l}\right) \tag{5-10}$$

Where Z_l is the feature map at layer l, and $W_{e,l}, b_{e,l}$ are the layer-specific weights and biases. Max pooling reduces dimensionality while retaining critical features. The decoder restores spatial resolution using up-sampling skip connections, ensuring the recovery of spatial details lost during encoding. The decoding operation is formulated as equation (5-16):

$$\hat{Y}_l = Z_l + g\left(W_{d,l} * \left[Z_l, Z_{\text{skip}}\right] + b_{d,l}\right) \tag{5-11}$$

Where Z_{skip} represents encoder features passed to the decoder. Skip connections enhance reconstruction accuracy by preventing feature degradation. Skip connections concatenate encoder and decoder feature maps to retain spatial information:

$$Z_{\text{skip}} = \text{Concat}(Z_{\text{encoder}}, Z_{\text{decoder}}) \tag{5-12}$$

This Equation improves training stability and enhances high-resolution reconstructions, making Residual U-Net effective for far-field acoustic prediction.

d) Discriminator:

The discriminator (Fig. 5-6) employs a hierarchical feature extraction process using convolutional layers to distinguish real and generated samples. The input feature map undergoes successive convolutional transformations, reducing spatial resolution while increasing the number of feature channels. Starting with a 512^2 input, a 3×3 convolution reduces it to 128^2 with 64 channels, followed by further down sampling to 64^2 with 256 channels. This enables the network to capture key spatial features for classification. A final fully connected layer predicts the probability of the input being real, ensuring effective adversarial training.

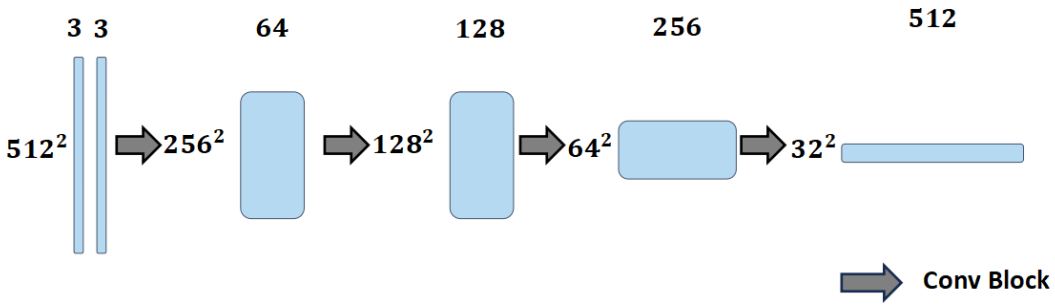


Fig. 5-6. Convolutional Feature Extraction Process.

The discriminator distinguishes between real and generated samples by extracting hierarchical features through convolutional layers and making binary decisions via a fully connected layer with a sigmoid activation [138]:

$$D(X_D;\theta_D)=\sigma(W_D*X_D+b_D) \tag{5-13}$$

Where X represents the input sample (either real or generated), $D(X_D)$ denotes the probability that X_D is a real sample, W_D,b_D are the trainable weights and biases, and $\sigma(\cdot)$ is the activation function, commonly Sigmoid in standard GANs or identity mapping in Wasserstein GANs (WGANs) [139]. Through this hierarchical feature extraction, the discriminator learns to accurately distinguish generated far-field distributions from real data, ensuring robust adversarial training.

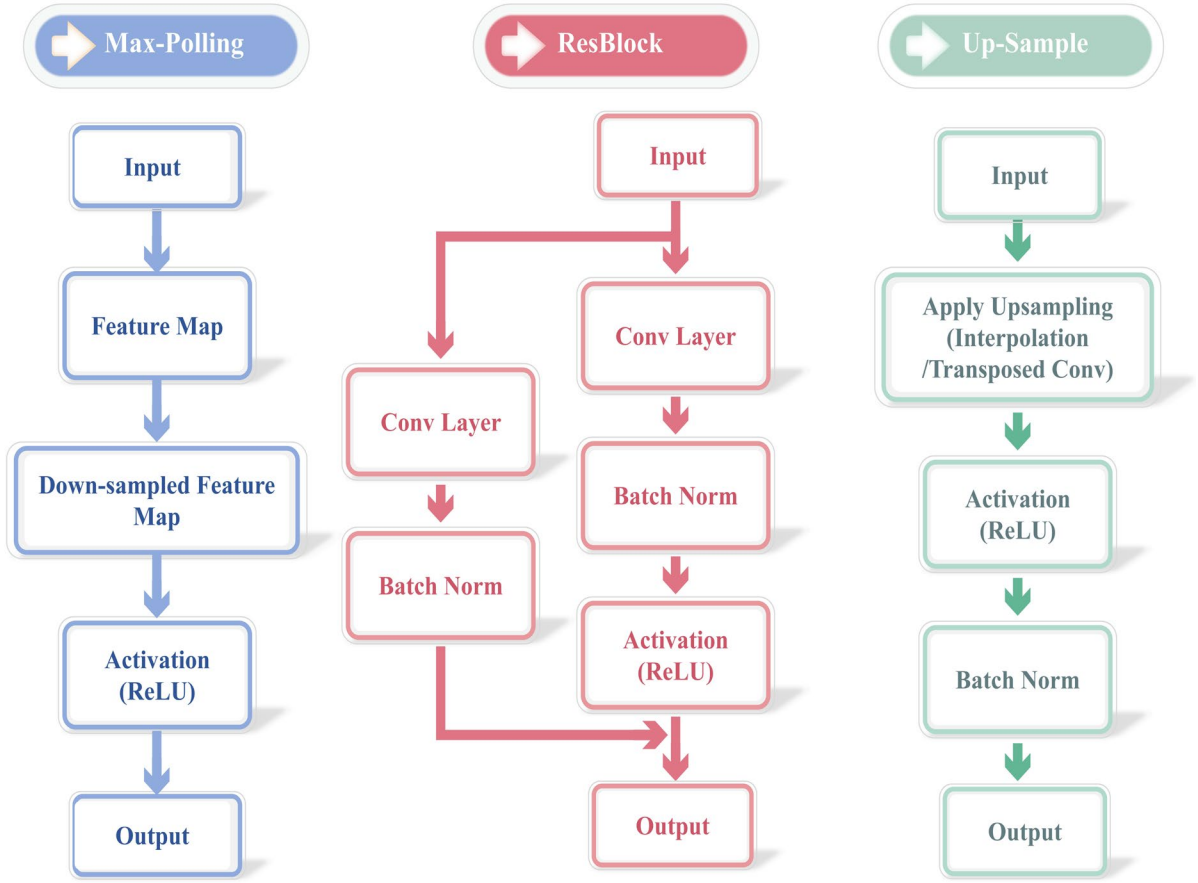


Fig. 5-7. Network Components for Generator and Discriminator.

In the generator and discriminator architecture, several fundamental building blocks contribute to extraction, transformation, and reconstruction. The Max Pooling module (Fig. 5-7) performs spatial downsampling by selecting the maximum value within non-overlapping local regions, effectively reducing resolution while preserving dominant features. This operation facilitates abstraction and controls computational complexity during feature encoding [140]. The ResBlock (Fig. 5-7) introduces residual connections, mitigating gradient vanishing issues and preserving spatial details, making it particularly effective in deep network structures such as U-Net and residual encoder-decoder architectures [141]. The Up-Sample module (Fig. 5-7) restores spatial resolution in the decoder, utilizing interpolation or transposed convolutions followed by activation and normalization [142]. These components collectively enhance the performance of the generator and discriminator, contributing to more stable adversarial training and improved reconstruction of 3D sound fields.

5.3 Evaluation and Results of GAN-Based Far-Field Directivity Prediction

5.3.1 Error Distribution Analysis

This section compares the performance of U-Net and Encoder-Decoder architectures in predicting the far-field directivity of a parametric array loudspeaker from near-field input. The evaluation includes visual comparisons with simulation results, error maps, and statistical error analysis.

Fig. 5-8 illustrates the comparison between the U-Net and Encoder-Decoder models in predicting the far-field directivity from near-field input data. The top section presents the input conditions and simulation ground truth, where a sample from the test dataset is selected for visualization. In this example, the steering angle is set to 20°, with eight elements in the array, forming a uniform array with a spacing of 1.25 cm. The middle and bottom sections correspond to the U-Net and Encoder-Decoder predictions, respectively, each consisting of three key components: the FEM solution (ground truth), the GAN model prediction, and the corresponding error map, which highlights the absolute error distribution between the predicted and ground-truth values.

The error maps in Fig. 5-8 on the rightmost column demonstrate that both models achieve a well-converged solution, enabling far-field directivity prediction with low errors. However, notable differences remain between the two models. The U-Net prediction more closely matches the FEM solution, displaying lower overall errors across the spatial domain. This can be attributed to the skip connections in U-Net, which help retain fine spatial features and minimize information loss during decoding. In contrast, the Encoder-Decoder model exhibits higher localized errors, particularly in regions with sharp gradient transitions, as evidenced by more intense variations in the error map. These results align with the statistical error analysis, further confirming that U-Net provides more stable and accurate directivity estimations compared to the Encoder-Decoder architecture.

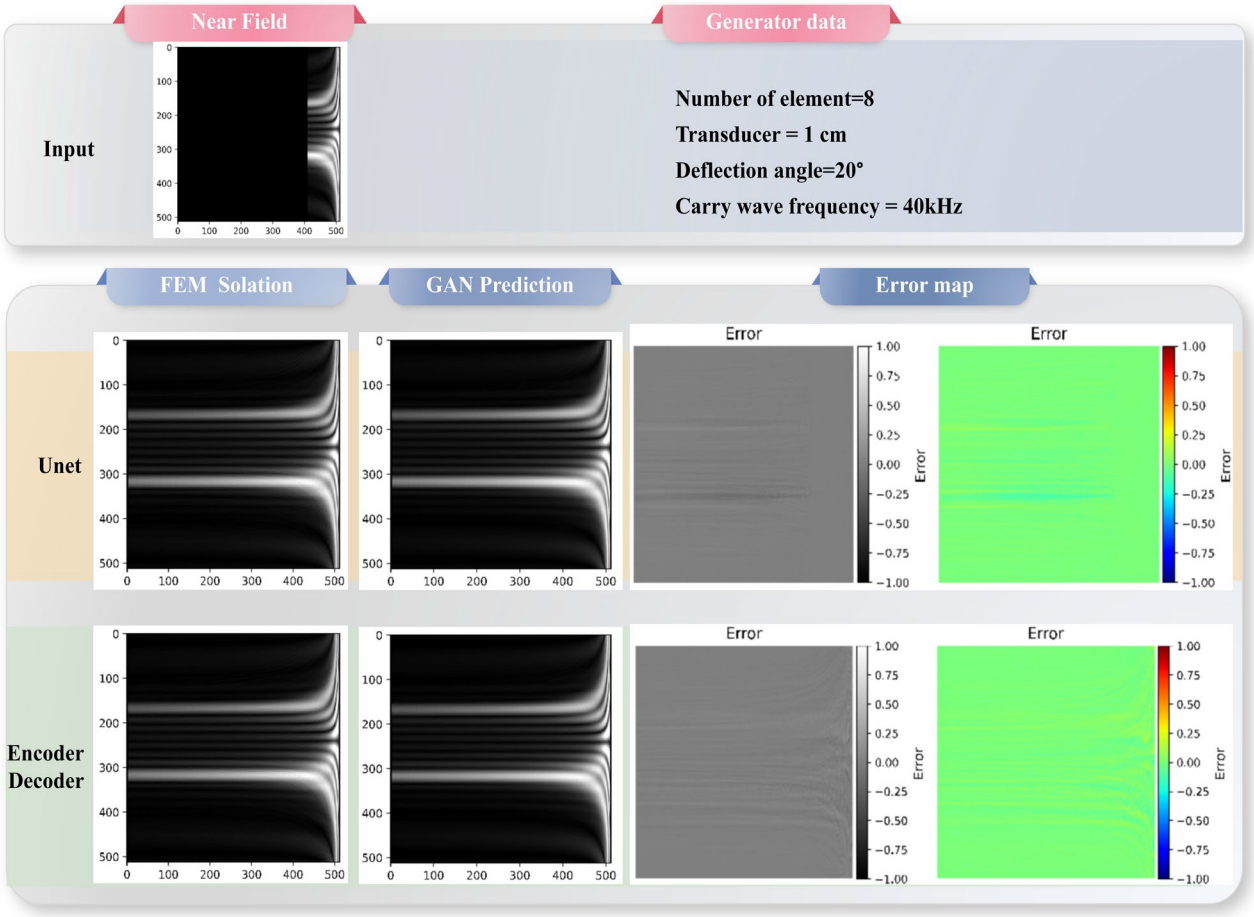


Fig. 5-8. Comparison of the estimation results from U-Net and Encoder-Decoder models. Ground truth (top), U-Net estimate (middle), Encoder–Decoder estimate (bottom); rightmost column shows absolute error.

Fig. 5-9 shows the error distribution for both models on the full test set of 350 samples. The Encoder–Decoder has a narrow distribution with a peak at lower error. The U-Net has a wider spread. The wider spread indicates higher variability in its predictions. Fig. 5-9(a) is the distribution of MAE, while Fig. 5-9(b) presents the corresponding distribution of root mean squared error (RMSE). In both subfigures, the histograms are overlaid with kernel density estimates (KDEs) to provide a smooth, continuous summary of each distribution; clear differences between the two architectures are evident in both the histograms and the KDE curves.

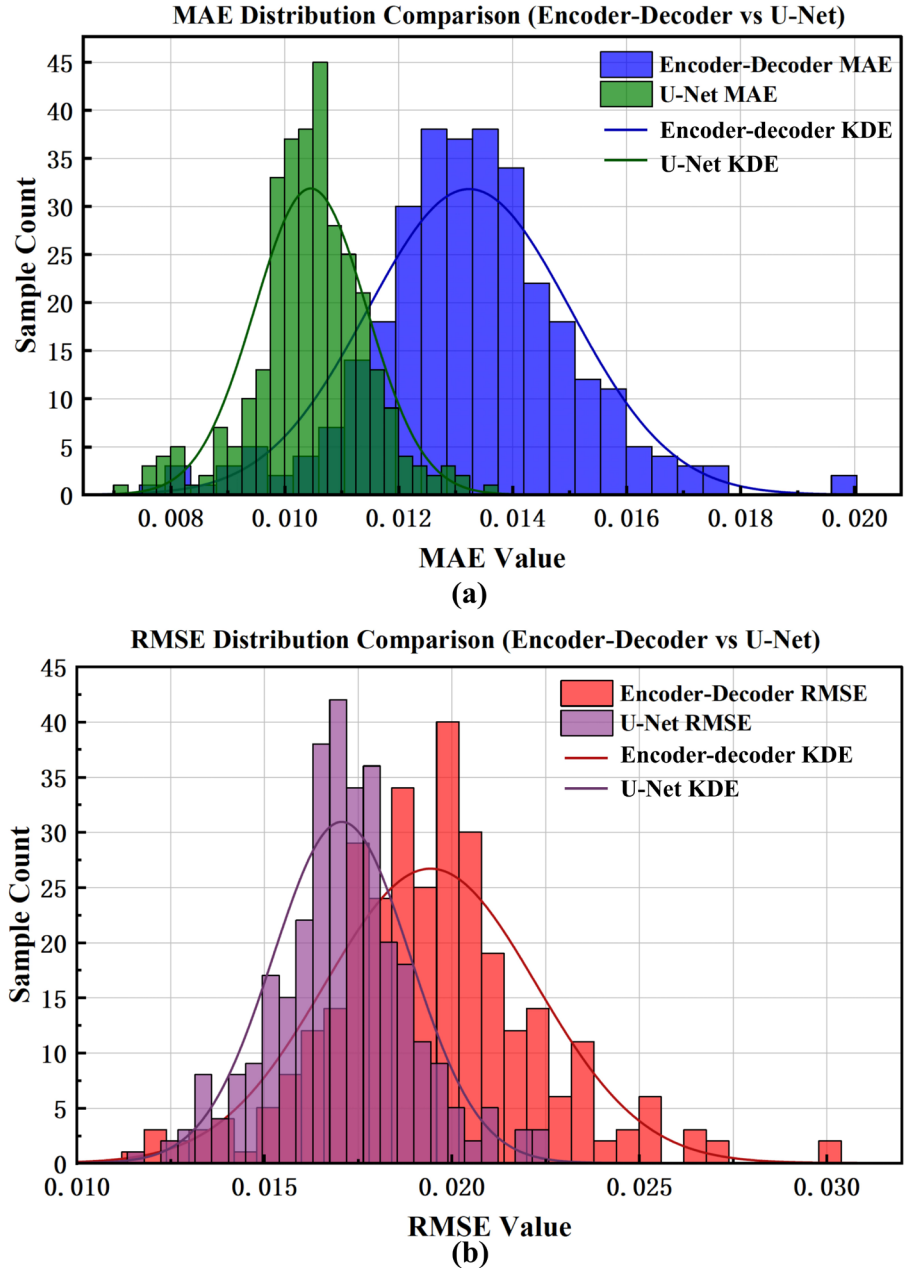


Fig. 5-9. MAE distribution(a) and RMSE distribution(b) of U-Net and Encoder-Decoder models. The histogram represents the frequency of different error magnitudes across the test dataset.

The results show that the U-Net model outperforms the Encoder–Decoder on both metrics. In numbers, U-Net reaches an average MAE of 0.0105, while Encoder–Decoder reaches 0.0132. U-Net also shows a tighter error distribution at lower values. This pattern indicates higher accuracy and greater stability across samples. A similar result appears for RMSE. U-Net has an average RMSE of 0.0171, while Encoder–Decoder has 0.0194. The concentration of RMSE

at lower magnitudes means U-Net reduces average error and also avoids extreme errors. As a result, U-Net is the more reliable model for directivity prediction.

In contrast, the Encoder–Decoder shows a broader error distribution with heavier tails for both MAE and RMSE. The wider spread reflects higher variability in prediction quality. The model is more sensitive to certain inputs and generalises less well across array layouts or steering angles. A likely reason is the lack of skip connections. The downsampling and upsampling lose important spatial features and reduce performance in complex cases..

5.3.2 Comparison of Far-Field Directivity Patterns

To validate the proposed predictive framework, far-field directivity measurements were performed using the experimental setup described in Chapter 4. A linear parametric array loudspeaker was evaluated at four steering angles: 60°, 65°, 70°, and 75°, under two carrier frequencies of 40 kHz and 42 kHz. The resulting SPL data, measured at a fixed distance across a 180° angular sector, are presented in Fig. 5-10. Subfigures (a)–(d) correspond to 40 kHz, while (e)–(h) present results for 42 kHz. These measurements were obtained using a uniform array configuration and are directly compared against full-wave FEM simulations. Because the GAN and FEM differ only slightly (RMSE = 0.0171; far-field deviation = 1.71%), for analysis we treat the FEM result as equivalent to the prediction and use it as the reference.

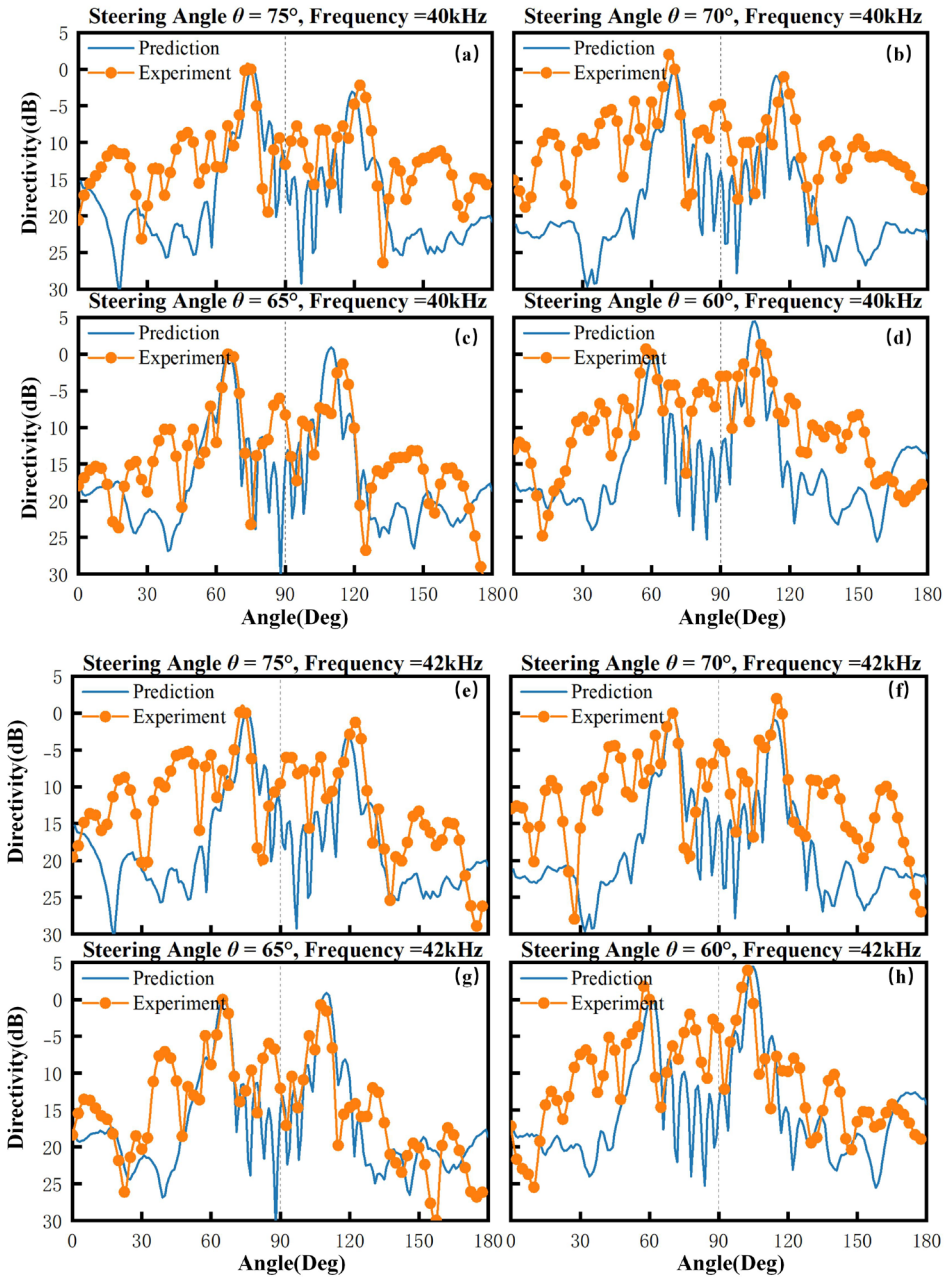


Fig. 5-10. Comparison of Experimental, FEM-Simulated, and GAN-Predicted Far-Field Directivity Patterns at Four Steering Angles (60°, 65°, 70°, 75°).

As shown in Fig. 5-10, the prediction result captures the far-field beam and the grating lobes at all tested steering angles and frequencies. The main lobes match the intended steering direction and align with the measurements (for example, around 75° in (a) and (e)). The positions and amplitudes of the secondary lobes also match well. These findings show that the model reproduces the key spatial features of the far-field pattern.

However, some differences appear in the sidelobe regions, especially where the field changes quickly or the SPL is low. The differences are larger near 0°–30° and 150°–180°. Likely causes include experimental noise, small transducer mismatches, and boundary reflections not modelled in the simulation. The 2.5° measurement step may also smooth fine structure predicted by the model.

Importantly, the GAN-predicted results exhibit almost identical trends to the FEM simulations across all tested configurations. Both models successfully capture the main lobe direction, grating lobe structures, and overall spatial distribution of the far-field directivity. The maximum point-wise deviation between the GAN and FEM output remains below 1%, confirming that the GAN model achieves high accuracy while significantly reducing computational cost. Quantitatively, the MAE and RMSE between FEM simulations and experimental measurements are approximately 7.7 dB and 9.3 dB, respectively. Despite minor mismatches inside lobe structures, these results indicate that the prediction framework robustly models the spatial characteristics of parametric loudspeaker arrays.

Overall, Fig. 5-10 validates the fidelity of both FEM and GAN models in reproducing measured acoustic fields. Their consistent performance across varying frequencies and steering angles supports their use as reliable tools for the prediction, analysis, and optimisation of directional acoustic systems.

5.4 Summary

This chapter presented a deep learning framework for efficient prediction of PAL far-field directivity, addressing the computational burden of full wave numerical simulation. The model maps near-field or configuration inputs to far-field, peak normalised dB directivity on matched angle grids, with physics informed losses to promote consistency with acoustic behaviour.

Tests cover steering from 60° to 75° at 40 kHz and 42 kHz. We can find that the main-lobe angles align with experiment. Sidelobe gaps appear at angular extremes due to noise, small phase errors, and the 2.5° measurement grid. On average, FEM vs. measurement gives MAE ≈ 7.7 dB and RMSE ≈ 9.3 dB, which sets a practical reference band for model–experiment agreement.

Two generator designs are studied in a GAN setting: U-Net and an Encoder–Decoder. U-Net reaches MAE = 0.0105 and RMSE = 0.0171. It outperforms a plain Encoder–Decoder (MAE = 0.0132, RMSE = 0.0194) and preserves fine spatial detail.

For efficiency, inference cuts per-pattern runtime from minutes with FEM to the sub-second range. This speed enables batch exploration of array layouts and rapid iteration. The framework is suitable as a design-time surrogate for far-field directivity studies and for adaptive beamforming.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This dissertation addressed far-field directivity prediction and array configuration for PALs under nonlinear propagation. We combined analytical modelling, numerical simulation, optimisation, and learning-based inference to improve accuracy, efficiency, and design practicality. Across measurement-referenced benchmarks, the proposed methods reduced directivity error and computation cost while preserving interpretability, enabling data-driven and physics-consistent PAL design. The main contributions of this study are summarized as follows:

In **Chapter 2**, **a detailed literature review is conducted to identify the limitations of existing analytical, numerical, and data-driven methods for modelling PAL directivity.** It is found that while formula-based models such as the Westervelt equation offer analytical insight, they often fall short in handling complex geometries and nonlinear effects. Numerical simulations, particularly FEM, provide high accuracy but are computationally expensive. Deep learning models emerge as promising alternatives, but challenges remain in their physical interpretability and data dependency. This review establishes the foundation and motivation for developing hybrid modelling frameworks that combine physical principles with machine learning.

In **Chapter 3**, **a dual convolution-based model is proposed to improve the prediction of far-field directivity and grating lobe behaviour.** This model leverages spatial impulse responses and element configuration data to characterize nonlinear propagation more accurately. Quantitatively, against peak-normalised measurements sampled from −90° to 90° (5° step), the dual-convolution approach reduces global RMSE from 45.23 dB (single-convolution) to 15.51 dB, and main-lobe (±20°) RMSE from 24.59 dB to 6.84 dB. It offering clearer insight into side lobe/grating-lobe formation than traditional formulations. Because the method relies on lightweight convolution operations rather than full-wave meshing, it functions as a fast, interpretable tool for early-stage array design and parameter sweeps.

In **Chapter 4**, **a phased array speaker optimisation framework is established based on PSO.** PAL array design as a multi-objective optimisation (grating-lobe suppression and main-lobe control) and validated PSO-optimised, non-uniform layouts with FEM and 2.0 m measurements (2.5° grid). Compared with uniform arrays, the optimised layouts consistently strengthen the main beam and suppress grating lobes in the mid-to-high steering range (60°–75°), achieving side lobe reductions up to 5 dB. At lower steering (60°), the advantage narrows, reflecting a practical trade-off between aggressive steering and beam robustness. Despite modest simulation–measurement gaps due to hardware tolerances and setup/environmental factors, the overall trends align, confirming the real-world applicability of the PSO-based design and providing practical guidance for deploying PALs under varied operating conditions.

In **Chapter 5**, **a deep learning-based prediction system is introduced, using Encoder-Decoder and GAN architectures to reconstruct far-field directivity patterns from sparse near-field measurements.** The GAN model recovers PAL far-field structure from sparse inputs with high fidelity and practical speed. Across the test set (peak-normalised dB directivity on matched angle grids), U-Net achieves MAE of 0.0105 and RMSE of 0.0171, outperforming a plain Encoder–Decoder and indicating better preservation of fine spatial detail; the GAN surrogate attains comparable errors while yielding sharper high-frequency features. Under measurement validation at 2.0 m with 2.5° resolution and steering angles 60°–75° at 40/42 kHz, GAN Model reproduces the steered main lobe and grating-lobe behaviour observed experimentally with typical single-digit-dB average errors. The GAN surrogate closely tracks FEM and measurements while cutting per-pattern runtime from minutes to the second range, enabling rapid, measurement-referenced prediction suitable for real-time or resource-constrained PAL design.

Overall, this research proposes an integrated framework for modelling, optimizing, and predicting PAL directivity by combining analytical methods, numerical simulations, machine learning, and experimental measurements, shortening design cycles from simulation-led iteration to rapid, measurement-referenced evaluation suitable for real deployments. The following achievements have been made in this study:

- A GAN-based far-field prediction framework that reconstructs patterns from sparse inputs with close agreement to FEM/experiment while cutting per-pattern runtime from minutes to the sub-second range for batch exploration.

- A PSO-based array configuration strategy that adapts to different steering angles and provides improved control over directivity and side lobe levels.
- A dual-convolution directivity model that fuses spatial impulse responses with array geometry, improving grating- and side lobe prediction and reducing measurement-referenced error for early-stage design.
- A systematic review and classification of PAL modelling methods that establishes a common, measurement-referenced evaluation protocol and motivates hybrid approaches balancing accuracy and runtime.

These contributions collectively provide a stronger foundation for accurate, efficient, and interpretable PAL design and control. They enable reproducible, measurement-referenced benchmarking, shorten design cycles from simulation-heavy iteration to rapid evaluation, and support actionable array synthesis under steering. In practice, this makes it more feasible to deploy PALs in targeted audio, automotive, and spatial-sound applications, while also charting a clear path toward 3D extensions, adaptive beamforming, and uncertainty-aware modelling in future work.

6.2 Future work

Based on the conclusions above and considering the limitations of the existing work, future research could be carried out in the following areas:

- **Extension to 3D Modelling and Complex Array Structures**
  All modelling and optimisation methods proposed in this study, including the dual convolution model, PSO-based array optimisation, and GAN-based prediction framework, are currently based on 2D simplifications. Future work should focus on extending these methods to 3D acoustic environments, which will more accurately represent real-world spatial propagation. In addition, the modelling framework should be adapted to accommodate more complex array geometries such as circular, planar, or curved arrays, and support various modulation schemes and emission patterns. This will enhance the flexibility and applicability of the proposed approaches in practical PAL applications.
- **Enhancement of Experimental Setup and Data Acquisition Efficiency**

The current experimental platform suffers from limitations in measurement speed and spatial resolution, leading to inefficiencies and increased susceptibility to error. Future studies should aim to develop an improved measurement system with automated scanning, higher spatial accuracy, and synchronized control. Such advancements would enable more reliable data collection for both training and validation of predictive models, especially in extended 3D sound fields.

- **Optimisation of GAN Performance for Sparse Input and Extended Distance Prediction**

  Although the GAN-based model in this study shows robust performance in predicting far-field directivity, its dependency on dense near-field data remains a limitation. Future efforts should be directed toward improving the model's ability to generate high-fidelity predictions with fewer input samples. This includes optimizing the network architecture for better generalization and extending the effective prediction range to cover broader far-field distances and angular spans, thus increasing its practical usability in real-time applications.

- **Integration of Physics-Informed Constraints into Deep Learning Models**

  The current GAN framework is almost data-driven, which just add the basic wave equation to limit the result. Future research should investigate the integration of physics-informed constraints, such as those derived from the Westervelt or KZK equations, into the training process. This can be achieved by introducing acoustic consistency terms into the loss function or applying differential equation-based regularizations. Such integration is expected to improve model interpretability, stability, and alignment with known nonlinear acoustic behaviour.

References

[1] D. Olszewski and K. Linhard, “3g-3 optimum array configuration for parametric ultrasound loudspeakers using standard emitters,” in *2006 IEEE Ultrasonics Symposium*, 2006.

[2] N. Tanaka and M. Tanaka, “Active noise control using a steerable parametric array loudspeaker,” *J. Acoust. Soc. Am.*, vol. 127, no. 6, pp. 3526–3537, 2010.

[3] J. Yang, W.-S. Gan, K.-S. Tan, and M.-H. Er, “Acoustic beamforming of a parametric speaker comprising ultrasonic transducers,” *Sens. Actuators Phys.*, vol. 125, no. 1, pp. 91–99, 2005.

[4] P. J. Westervelt, “Parametric acoustic array,” *J. Acoust. Soc. Am.*, vol. 35, no. 4, pp. 535–537, 1963.

[5] M. Cervenka and M. Bednarik, “Non-paraxial model for a parametric acoustic array,” *J. Acoust. Soc. Am.*, vol. 134, no. 2, pp. 933–938, 2013.

[6] R. Pallavi Reddy, S. Srikeerthi, C. Hema, and B. Sneha Latha, “Enhancing Audio Synthesis with WAVEGAN: A Generative Adversarial Network (GAN) Approach,” in *Cognitive Science and Technology*, Singapore: Springer Nature Singapore, 2025, pp. 107–113.

[7] C. Donahue, J. McAuley, and M. Puckette, “Adversarial Audio Synthesis.” arXiv, Feb. 09, 2019.

[8] J. Zhong and X. Qiu, *Acoustic Waves Generated by Parametric Array Loudspeakers*. CRC Press, 2024.

[9] K. Tanaka, C. Shi, and Y. Kajikawa, “Binaural active noise control using parametric array loudspeakers,” *Appl. Acoust.*, vol. 116, pp. 170–176, 2017.

[10] S. T. Sasi, A. V. Nath, and A. V. R, “Directional Steering of Audio Signal Using Parametric Array in Air,” *Int. J. Eng. Res. Technol.*, vol. 3, no. 1, Jan. 2014.

[11] A. Ritty, “Directional loudspeaker using a parametric array,” *Acta Polytech.*, vol. 46, no. 4, 2006.

[12] Y. Nishiura, M. Hoshiyama, and Y. Konagaya, “Use of parametric speaker for older people with dementia in a residential care setting: A preliminary study of two cases,” *Hong Kong J. Occup. Ther.*, vol. 31, no. 1, pp. 30–35, June 2018.

[13] A. Burka, A. Qin, and D. D. Lee, “An application of parametric speaker technology to bus-pedestrian collision warning,” in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2014, pp. 948–953.

[14] S. Kashiwase and K. Kondo, “Towards a parametric speaker system with human head tracking beam control,” in *2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE)*, 2014, pp. 22–23.

[15] A. Adachi and H. Hashiguchi, “Application of parametric speakers to radio acoustic sounding system,” *Atmospheric Meas. Tech.*, vol. 12, no. 10, pp. 5699–5715, 2019.

[16] T. Komatsuzaki and Y. Iwata, “Active noise control using high-directional parametric loudspeaker,” *J. Environ. Eng.*, vol. 6, no. 1, pp. 140–149, 2011.

[17] H. Furuhashi, Y. Mori, and S. Suzuki, “Active noise control of a plane sound wave by a parametric speaker,” in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 2019, vol. 260, no. 1, pp. 227–233.

[18] J. Zhong, R. Kirby, and X. Qiu, “A spherical expansion for audio sounds generated by a circular parametric array loudspeaker,” *J. Acoust. Soc. Am.*, vol. 147, no. 5, pp. 3502–3510, 2020.

[19] P. Ji, C. Ye, and J. Tian, “Development of a directional loudspeaker system for sound reproduction,” in *2007 IEEE International Conference on Multimedia and Expo*, 2007, pp. 591–594.

[20] M. B. Moffett and R. H. Mellen, “Model for parametric acoustic sources,” *J. Acoust. Soc. Am.*, vol. 61, no. 2, pp. 325–337, 1977.

[21] W.-S. Gan, J. Yang, and T. Kamakura, “A review of parametric acoustic array in air,” *Appl. Acoust.*, vol. 73, no. 12, pp. 1211–1219, 2012.

[22] E. M. Ballad, S. Y. Vezirov, K. Pfleiderer, I. Y. Solodov, and G. Busse, “Nonlinear modulation technique for NDE with air-coupled ultrasound,” *Ultrasonics*, vol. 42, no. 1–9, pp. 1031–1036, 2004.

[23] A. Loriga, “Nonlinear Model and its Inverse of an Audio System.” arXiv, Feb. 28, 2017.

[24] F. J. Pompei, “Parametric audio system,” June. 24, 2008.

[25] W. J. Lamb and R. M. Aarts, “Driving of parametric loudspeakers,” July. 07, 2015.

[26] B. P. Lathi and Z. Ding, *Modern digital and analog communication systems*, vol. 3. Oxford university press New York, 1998.

[27] K. H. Hamza and D. Nirmal, “A review of GaN HEMT broadband power amplifiers,” *AEU-Int. J. Electron. Commun.*, vol. 116, p. 153040, 2020.

[28] W.-S. Gan, P. Ji, E. L. Tan, and J. Yang, “Distortion analysis and reduction for the parametric array,” in *Audio Engineering Society Convention 124*, 2008.

[29] M. Yoneyama, J. Fujimoto, Y. Kawamo, and S. Sasabe, “The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design,” *J. Acoust. Soc. Am.*, vol. 73, no. 5, pp. 1532–1536, 1983.

[30] E.-L. Tan, P. Ji, and W.-S. Gan, “On preprocessing techniques for bandlimited parametric loudspeakers,” *Appl. Acoust.*, vol. 71, no. 5, pp. 486–492, 2010.

[31] T. Kamakura, “Development of parametric loudspeaker for practical use,” in *Proc. 10th Int. Symp. Nonlinear Acoustics*, 1984, pp. 147–150.

[32] P. Ji, W.-S. Gan, E.-L. Tan, and J. Yang, “Performance analysis on recursive single-sideband amplitude modulation for parametric loudspeakers,” in *2010 IEEE International Conference on Multimedia and Expo*, 2010, pp. 748–753.

[33] S.-H. Yu, Y.-F. Hsieh, P.-Y. Lai, Y.-L. Chen, C.-P. Yang, and K. Lin, “FPGA-based resonant-frequency-tracking power amplifier for ultrasonic transducer,” in *2015 International Conference on Applied Electronics (AE)*, 2015, pp. 285–288.

[34] J. Mikulka, D. Hladkỳ, and J. Sliž, “Parametric array as a source of audible signal,” in *2016 Progress in Electromagnetic Research Symposium (PIERS)*, 2016, pp. 3610–3614.

[35] M. Arnela, O. Guasch, P. Sánchez-Martín, J. Camps, R. M. Alsina-Pagès, and C. Martínez-Suquía, “Construction of an omnidirectional parametric loudspeaker consisting in a spherical distribution of ultrasound transducers,” *sensors*, vol. 18, no. 12, p. 4317, 2018.

[36] 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, 1988.

[37] S. Lin, Y. Chen, and Z. J. Wong, “High-performance optical beam steering with nanophotonics,” *Nanophotonics*, vol. 11, no. 11, pp. 2617–2638, May 2022.

[38] J. C. Chen, K. Yao, and R. E. Hudson, “Source localization and beamforming,” *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 30–39, 2002.

[39] Y. Wang, X. Li, L. Xu, and L. Xu, “SSB modulation of the ultrasonic carrier for a parametric loudspeaker,” in *2009 International Conference on Electronic Computer Technology*, 2009, pp. 669–673.

[40] E. Gaalaas, “Class d audio amplifiers: What, why, and how,” *Analog Dialogue*, vol. 40, no. 6, pp. 1–7, 2006.

[41] G. Breed and E. Director, “An overview of common techniques for power amplifier linearization,” *IEEE Microw. Wirel. Comp Lett*, vol. 18, 2008.

[42] Y. Kang, T. Ge, H. He, and J. S. Chang, “A review of audio class D amplifiers,” in *2016 International Symposium on Integrated Circuits (ISIC)*, 2016, pp. 1–4.

[43] S. Kwon et al., “A 0.028% THD+ N, 91% power-efficiency, 3-level PWM Class-D amplifier with a true differential front-end,” in *2012 IEEE International Solid-State Circuits Conference*, 2012, pp. 96–98.

[44] X. Jiang, “Fundamentals of audio class D amplifier design: A review of schemes and architectures,” *IEEE Solid-State Circuits Mag.*, vol. 9, no. 3, pp. 14–25, 2017.

[45] M. K. Kazimierczuk, *RF power amplifiers*. John Wiley & Sons, 2014.

[46] B. Razavi, “RF Microelectronics,” *NJ USA*, 2011.

[47] H. Choi, “Prelinearized class-b power amplifier for piezoelectric transducers and portable ultrasound systems,” *Sensors*, vol. 19, no. 2, p. 287, 2019.

[48] A. Katz, “Linearization: Reducing distortion in power amplifiers,” *IEEE Microw. Mag.*, vol. 2, no. 4, pp. 37–49, 2001.

[49] H. Choi, “Development of a class-c power amplifier with diode expander architecture for point-of-care ultrasound systems,” *Micromachines*, vol. 10, no. 10, p. 697, 2019.

[50] G. Velasco-Quesada, M. Román-Lumbreras, R. Pérez-Delgado, and A. Conesa-Roca, “Class H power amplifier for power saving in fluxgate current transducers,” *IEEE Sens. J.*, vol. 16, no. 8, pp. 2322–2330, 2016.

[51] T. Piessens and M. Steyaert, “Highly efficient xDSL line drivers in 0.35-µm CMOS using a self-oscillating power amplifier,” *IEEE J. Solid-State Circuits*, vol. 38, no. 1, pp. 22–29, 2003.

[52] P. Blomley, “New approach to Class B amplifier design,” *Wirel. World*, pp. 57–61, 1971.

[53] M. Mauerer, A. Tüysüz, and J. W. Kolar, “Distortion analysis of low-THD/high-bandwidth GaN/SiC class-D amplifier power stages,” in *2015 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2015, pp. 2563–2571.

[54] P. P. G. Microsemi, “Gallium nitride (GaN) versus silicon carbide (SiC) In the high frequency (RF) and power switching applications,” *Digi-Key*, vol. 82, 2014.

[55] T. Ueda, “Recent advances and future prospects on GaN-based power devices,” in *2014 International Power Electronics Conference (ipec-hiroshima 2014-ECCE ASIA)*, 2014, pp. 2075–2078.

[56] M. Mauerer and J. W. Kolar, “Distortion minimization for ultra-low THD class-D power amplifiers,” *CPSS Trans. Power Electron. Appl.*, vol. 3, no. 4, pp. 324–338, 2018.

[57] J. Sangid, G. Long, P. Mitchell, B. J. Blalock, D. J. Costinett, and L. M. Tolbert, “Comparison of 60V GaN and Si devices for class D audio applications,” in *2018 IEEE 6th Workshop on Wide Bandgap Power Devices and Applications (wipda)*, 2018, pp. 73–76.

[58] J. Chung, R. McKenzie, and W. T. Ng, “A comparison between GaN and silicon based class D audio power amplifiers with pulse density modulation,” in *2016 13th IEEE International Conference on Solid-state and Integrated Circuit Technology (ICSICT)*, 2016, pp. 90–93.

[59] L. Gu et al., “Multiphase GaN class-D resonant amplifier for high-intensity focused ultrasound,” in *2019 20th Workshop on Control and Modeling for Power Electronics (COMPEL)*, 2019, pp. 1–6.

[60] J. Kuroda and Y. Oikawa, “Parametric speaker consisting of small number of transducers with sonic crystal waveguide,” *Acoust. Sci. Technol.*, vol. 41, no. 6, pp. 865–876, 2020.

[61] Y. Je, H. Lee, and W. Moon, “The impact of micromachined ultrasonic radiators on the efficiency of transducers in air,” *Ultrasonics*, vol. 53, no. 6, pp. 1124–1134, 2013.

[62] Z. Shao, S. Pala, Y. Liang, Y. Peng, and L. Lin, “A single chip directional loudspeaker based on PMUTS,” in *2021 IEEE 34th International Conference on Micro Electro Mechanical Systems (MEMS)*, 2021, pp. 895–898.

[63] Y. Je, H. Lee, K. Been, and W. Moon, “A micromachined efficient parametric array loudspeaker with a wide radiation frequency band,” *J. Acoust. Soc. Am.*, vol. 137, no. 4, pp. 1732–1743, 2015.

[64] C. Hedberg and H. Gazisaeidi, “Variation of high power air transducer,” vol. Nov. 15–17, pp. 199–200, Jan. 2006.

[65] J.-Y. Kim, S.-S. Choi, S.-M. Song, I.-D. Kim, and W. Moon, “Design of high-efficiency power amplifier system for high-directional speaker,” in *2016 IEEE 8th International Power Electronics and Motion Control Conference (IPEMC-ECCE Asia)*, 2016, pp. 1093–1097.

[66] D. Florencio and Z. Zhang, “Driving parametric speakers as a function of tracked user location,” Dec. 31, 2015.

[67] C. Shi, Y. Kajikawa, and W.-S. Gan, “An overview of directivity control methods of the parametric array loudspeaker,” *APSIPA Trans. Signal Inf. Process.*, vol. 3, p. e20, 2014.

[68] H. Gao, J. Zheng, B. Qian, and X. Chen, “Near-field focusing emulation of near-field-focused arrays using far-field-focused arrays,” in *2024 Photonics & Electromagnetics Research Symposium (PIERS)*, 2024, pp. 1–5.

[69] C. Rowell and B. Derat, “Near-field measurement system and method for obtaining far-field characteristics,” July. 28, 2020.

[70] K. G. Foote, “Discriminating between the nearfield and the farfield of acoustic transducers,” *J. Acoust. Soc. Am.*, vol. 136, no. 4, pp. 1511–1517, 2014.

[71] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of acoustics*. John wiley & sons, 2000.

[72] K. T. Rich and T. D. Mast, “Methods to calibrate the absolute receive sensitivity of single-element, focused transducers,” *J. Acoust. Soc. Am.*, vol. 138, no. 3, pp. EL193–EL198, 2015.

[73] J. Zhong, R. Kirby, and X. Qiu, “The near field, westervelt far field, and inverse-law far field of the audio sound generated by parametric array loudspeakers,” *J. Acoust. Soc. Am.*, vol. 149, no. 3, pp. 1524–1535, 2021.

[74] I. S. O. Acoustics, “Attenuation of sound during propagation outdoors,” *Markham ISO*, vol. 533, pp. 1–8, 1996.

[75] H. E. Bass, L. C. Sutherland, and A. J. Zuckerwar, “Atmospheric absorption of sound: update,” 1990.

[76] H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. M. Hester, “Erratum: atmospheric absorption of sound: further developments [j. acoust. soc. am. 97, 680–683 (1995)],” *J. Acoust. Soc. Am.*, vol. 99, no. 2, pp. 1259–1259, 1996.

[77] “On sound generated aerodynamically I. General theory,” *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.*, vol. 211, no. 1107, pp. 564–587, Mar. 1952.

[78] E. A. Zabolotskaya, “Quasi-plane waves in the nonlinear acoustics of confined beams,” *Sov Phys Acoust*, vol. 15, no. 1, pp. 35–40, 1969.

[79] S. I. Aanonsen, T. Barkve, J. N. Tjo/tta, and S. Tjo/tta, “Distortion and harmonic generation in the nearfield of a finite amplitude sound beam,” *J. Acoust. Soc. Am.*, vol. 75, no. 3, pp. 749–768, 1984.

[80] D. T. Blackstock and M. F. Hamilton, *Nonlinear acoustics*. Academic Press, 1998.

[81] D. T. Blackstock, *Fundamentals of physical acoustics*. John Wiley & Sons, 2000.

[82] A. Rozanova, “The khokhlov–zabolotskaya–kuznetsov equation,” *Comptes Rendus Mathématique*, vol. 344, no. 5, pp. 337–342, 2007.

[83] J. M. Burgers, “A mathematical model illustrating the theory of turbulence,” *Adv. Appl. Mech.*, vol. 1, pp. 171–199, 1948.

[84] J. Zhong, H. Zou, J. Lu, and D. Zhang, “A modified convolution model for calculating the far field directivity of a parametric array loudspeaker,” *J. Acoust. Soc. Am.*, vol. 153, no. 3, pp. 1439–1451, 2023.

[85] V. P. Kuznetsov, “Equation of nonlinear acoustics,” *Sov Phys Acoust*, vol. 16, pp. 467–470, 1971.

[86] S. Xu, L. Qin, G. Yang, and K. Zhang, “Optimization of sparse planar arrays using genetic algorithms,” in *2023 International Applied Computational Electromagnetics Society Symposium (aces-china)*, 2023, pp. 1–3.

[87] M. Saleem, Z. U. Rehman, A. Sadiq, U. Zahoor, and M. R. Anjum, “Study of beamforming methods of phased array antenna with steering angle,” in *2016 Sixth International Conference on Innovative Computing Technology (INTECH)*, 2016, pp. 617–620.

[88] S. H. Wong, M. Kupnik, K. Butts-Pauly, and B. T. Khuri-Yakub, “P1B-10 advantages of capacitive micromachined ultrasonics transducers (CMUTs) for high intensity focused ultrasound (HIFU),” in *2007 IEEE Ultrasonics Symposium Proceedings*, 2007, pp. 1313–1316.

[89] K. Okita, K. Ono, S. Takagi, and Y. Matsumoto, “Development of high intensity focused ultrasound simulator for large-scale computing,” *Int. J. Numer. Methods Fluids*, vol. 65, no. 1–3, pp. 43–66, Jan. 2011.

[90] G. Taraldsen, “A generalized westervelt equation for nonlinear medical ultrasound,” *J. Acoust. Soc. Am.*, vol. 109, no. 4, pp. 1329–1333, 2001.

[91] Z. P. Izgi, P. O. Mohammed, R. P. Agarwal, M. A. Yousif, A. A. Lupas, and M. Abdelwahed, “Efficient study on westervelt-type equations to design metamaterials via symmetry analysis,” *Mathematics*, vol. 12, no. 18, p. 2855, 2024.

[92] M. R. Bai, C.-W. Tung, and C.-C. Lee, “Optimal design of loudspeaker arrays for robust cross-talk cancellation using the taguchi method and the genetic algorithm,” *J. Acoust. Soc. Am.*, vol. 117, no. 5, pp. 2802–2813, 2005.

[93] S. Adavanne, A. Politis, J. Nikunen, and T. Virtanen, “Sound event localization and detection of overlapping sources using convolutional recurrent neural networks,” *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 1, pp. 34–48, 2018.

[94] S. Pascual, A. Bonafonte, and J. Serrà, “SEGAN: speech enhancement generative adversarial network.” arXiv, June. 09, 2017.

[95] J. Wu, C. Zhang, T. Xue, B. Freeman, and J. Tenenbaum, "Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling," *Adv. Neural Inf. Process. Syst.*, vol. 29, 2016.

[96] T. Okuzono, T. Yoshida, and K. Sakagami, "Efficiency of room acoustic simulations with time-domain FEM including frequency-dependent absorbing boundary conditions: comparison with frequency-domain FEM," *Appl. Acoust.*, vol. 182, p. 108212, Nov. 2021.

[97] J. Saarelma, "Finite-difference time-domain solver for room acoustics using graphics processing units," *Masters Thesis Aalto Univ. Finl.*, 2013.

[98] M. Ochmann, "Boundary element acoustics fundamentals and computer codes." Acoustical Society of America, 2002.

[99] R. K. Hansen and P. A. Andersen, "The application of real time 3D acoustical imaging," in *IEEE Oceanic Engineering Society. OCEANS'98. Conference Proceedings (cat. No. 98CH36259)*, 1998, vol. 2, pp. 738–741.

[100] Z. Wan et al., "A review of acoustic devices based on suspended 2D materials and their composites," *Adv. Funct. Mater.*, vol. 34, no. 3, p. 2303519, Jan. 2024.

[101] C. Shi and Y. Kajikawa, "A convolution model for computing the far-field directivity of a parametric loudspeaker array," *J. Acoust. Soc. Am.*, vol. 137, no. 2, pp. 777–784, 2015.

[102] L. W. Schmmer, "Fundamentals of ultrasonic phased arrays," *Mod. Phys. Lett. B*, vol. 22, no. 11, pp. 917–921, May 2008.

[103] J. Zhong, R. Kirby, M. Karimi, and H. Zou, "A cylindrical expansion of the audio sound for a steerable parametric array loudspeaker," *J. Acoust. Soc. Am.*, vol. 150, no. 5, pp. 3797–3806, 2021.

[104] T. Zhuang, J.-X. Zhong, G. Ma, F. Niu, and J. Lu, "Modeling a phased array of parametric array loudspeakers using the spherical wave expansion," *IEEE Trans. Audio Speech Lang. Process.*, 2025.

[105] P. Dufossé, C. Enderli, L. Savy, and N. Hansen, "Phased-array antenna pattern optimization with evolution strategies," in *2020 IEEE Radar Conference (radarconf20)*, 2020, pp. 1–6.

[106] Y. Mu, P. Ji, W. Ji, M. Wu, and J. Yang, “Modeling and compensation for the distortion of parametric loudspeakers using a one-dimension volterra filter,” *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 12, pp. 2169–2181, 2014.

[107] S.-C. Wooh and Y. Shi, “Three-dimensional beam directivity of phase-steered ultrasound,” *J. Acoust. Soc. Am.*, vol. 105, no. 6, pp. 3275–3282, 1999.

[108] C. M. U. Guide, “COMSOL: burlington,” *Ma Usa*, 2018.

[109] R. Poli, “Analysis of the publications on the applications of particle swarm optimisation,” *J. Artif. Evol. Appl.*, vol. 2008, no. 1, p. 685175, Jan. 2008.

[110] Y. Shi, “Particle swarm optimization: developments, applications and resources,” in *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No. 01TH8546)*, 2001, vol. 1, pp. 81–86.

[111] V. Černý, “Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm,” *J. Optim. Theory Appl.*, vol. 45, no. 1, pp. 41–51, Jan. 1985.

[112] L. Ingber, “Simulated annealing: practice versus theory,” *Math. Comput. Model.*, vol. 18, no. 11, pp. 29–57, 1993.

[113] D. Henderson, S. H. Jacobson, and A. W. Johnson, “The theory and practice of simulated annealing,” in *Handbook of Metaheuristics*, vol. 57, F. Glover and G. A. Kochenberger, Eds. Boston: Kluwer Academic Publishers, 2003, pp. 287–319.

[114] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of Icnn’95-international Conference on Neural Networks*, 1995, vol. 4, pp. 1942–1948.

[115] D. E. Goldberg, “Genetic Algorithm in Search, Optimization and Machine Learning, Addison,” *W Esley Publ. Co. R Eading MA*, vol. 1, no. 98, p. 9, 1989.

[116] S. S.N. and D. S.N., *Introduction to genetic algorithms*. Berlin, Heidelberg: Springer, 2008.

[117] M. Clerc and J. Kennedy, “The particle swarm - explosion, stability, and convergence in a multidimensional complex space,” *IEEE Trans. Evol. Comput.*, vol. 6, no. 1, pp. 58–73, Feb. 2002.

[118] S. Kirkpatrick, J. C. D. Gelatt, and M. P. Vecchi, “Optimization by simulated annealing,” *Science*, May 1983.

[119] L. M. Rios and N. V. Sahinidis, “Derivative-free optimization: a review of algorithms and comparison of software implementations,” *J. Glob. Optim.*, vol. 56, no. 3, pp. 1247–1293, July 2013.

[120] T. Cox and P. d’Antonio, *Acoustic absorbers and diffusers: theory, design and application*. CRC press, 2016.

[121] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134.

[122] X. Karakonstantis and E. Fernandez-Grande, “Generative adversarial networks with physical sound field priors.” arXiv, Aug. 01, 2023.

[123] O. Ronneberger, P. Fischer, and T. Brox, “U-net: convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[124] V. Badrinarayanan, A. Handa, and R. Cipolla, “SegNet: a deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling.” arXiv, May. 27, 2015.

[125] W. Mallik, R. K. Jaiman, and J. Jelovica, “Predicting transmission loss in underwater acoustics using convolutional recurrent autoencoder network,” *J. Acoust. Soc. Am.*, vol. 152, no. 3, p. 1627, Sept. 2022.

[126] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation.” arXiv, Mar. 08, 2015.

[127] A. Vaswani et al., “Attention is all you need.” arXiv, Aug. 02, 2023.

[128] A. Dosovitskiy et al., “An image is worth 16x16 words: transformers for image recognition at scale.” arXiv, June. 03, 2021.

[129] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Adv. Neural Inf. Process. Syst.*, vol. 25, 2012.

[130] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *J. Big Data*, vol. 6, no. 1, p. 60, Dec. 2019.

[131] I. Goodfellow et al., “Generative adversarial networks,” *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020.

[132] C. Ledig et al., “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4681–4690.

[133] I. Kavalerov, W. Czaja, and R. Chellappa, “A multi-class hinge loss for conditional gans,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 1290–1299.

[134] E. Schonfeld, B. Schiele, and A. Khoreva, “A u-net based discriminator for generative adversarial networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8207–8216.

[135] I. J. Goodfellow et al., “Generative adversarial nets,” *Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.

[136] F. Isensee et al., “nnU-net: self-adapting framework for U-net-based medical image segmentation.” arXiv, Sept. 27, 2018.

[137] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, no. 5786, pp. 504–507, July 2006.

[138] R. K. Srivastava, K. Greff, and J. Schmidhuber, “Highway networks.” arXiv, Nov. 03, 2015.

[139] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *Computer Vision – ECCV 2014*, vol. 8689, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 818–833.

[140] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[141] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.

[142] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv, Jan. 07, 2016.

END