Model-Driven Deep Neural Network for Enhanced AoA Estimation Using 5G gNB

Shengheng Liu[1, 2][*]**, Xingkang Li**[1]**, Zihuan Mao**[1, 2]**, Peng Liu**[2]**, and Yongming Huang**[1, 2][*]

[1]National Mobile Communications Research Laboratory, Southeast University, Nanjing, China
[2]Purple Mountain Laboratories, Nanjing, China
{s.liu; huangym}@seu.edu.cn

Abstract

High-accuracy positioning has become a fundamental enabler for intelligent connected devices. Nevertheless, the present wireless networks still rely on model-driven approaches to achieve positioning functionality, which are susceptible to performance degradation in practical scenarios, primarily due to hardware impairments. Integrating artificial intelligence into the positioning framework presents a promising solution to revolutionize the accuracy and robustness of location-based services. In this study, we address this challenge by reformulating the problem of angle-of-arrival (AoA) estimation into image reconstruction of spatial spectrum. To this end, we design a model-driven deep neural network (MoD-DNN), which can automatically calibrate the angular-dependent phase error. The proposed MoD-DNN approach employs an iterative optimization scheme between a convolutional neural network and a sparse conjugate gradient algorithm. Simulation and experimental results are presented to demonstrate the effectiveness of the proposed method in enhancing spectrum calibration and AoA estimation.

Introduction

Development of the forthcoming wireless communications entails the exploration and integration of cutting-edge technologies and paradigms for diverse and emerging applications (You et al. 2023). Among the essential components, high-accuracy positioning functionality assumes a critical role in guaranteeing a seamless user experience and accommodating various use cases such as emergency response, asset tracking, and location-based services (Sun, Melamed, and Kitani 2021). Despite being in its nascent stages, an escalating number of 5G mobile base stations (a.k.a. gNBs) have been deployed on a massive scale in both indoor and outdoor areas, offering an extraordinary platform for supporting positioning services (Pai and Kumar 2019). As a result, there is a renewed drive for the perception of the surrounding environment and the provision of location-aware services from the physical layer to the application layer.

In Release 16 of the 3GPP, which was finalized in 2021, several positioning schemes for 5G new radio (NR) have been introduced. These include Downlink-time-difference-of-arrival (DL-TDoA), uplink-time-difference-of-arrival (UL-TDoA), downlink-angle-of-departure (DL-AoD), uplink-angle-of-arrival (UL-AoA), and multi-round trip time (Multi-RTT) positioning. Among these schemes, UL-AoA has been identified as more feasible for mitigating the multipath effect of the receive signal and enhancing the positioning performance, as it aligns with the trend towards increasing array size and bandwidth. However, purely model-driven AoA estimation algorithms usually suffer from hardware impairments that can affect the array response used in the estimation process (Pan et al. 2022).

By leveraging artificial intelligence (AI) algorithms, such as machine learning and deep neural networks (DNN), wireless networks can now dynamically learn and adapt to real-world conditions, compensating for hardware impairments and environmental variations (Dai et al. 2022; Lee et al. 2023). To address this issue, data-driven methods, such as DNN-based direction finding techniques, have surged exponential popularity (Ghourchian, Allegue-Martinez, and Precup 2017; Mo and Morgado 2023). Nonetheless, the accuracy of such methods is heavily reliant on the availability of massive data sets, which hinders their translation to practical uses. The integration of neural network (NN) and model-driven methods (Zhu et al. 2022; Huang et al. 2023) provides inspiration for developing a model-driven deep learning method to mitigate the impact of hardware impairments and improve the AoA estimation performance.

Building on these recent advancements, this paper presents an innovative model-driven deep neural network (MoD-DNN) framework, featuring spectrum calibration capabilities to enhance AoA estimation. Simulation and experimental results demonstrate that the proposed framework yields superior AoA estimation performance compared to both purely model-driven and data-driven methods. To the best of our knowledge, this study marks the first AI-empowered positioning endeavor utilizing commodity 5G NR gNB. The key contributions of this paper include: (1) The problem formulation of an inverse problem from the coarray spatial spectrum (CSS) to a sparse solution of the over-completed AoA set; (2) Employing a one-dimensional convolutional neural network (1D-CNN) for CSS calibration, followed by a sparse conjugate gradient (SCG) algorithm for sparse AoA spectrum recovery; and (3) Proposing an iterative optimization strategy utilizing shared CNN coefficients between the CNN and SCG modules to establish

the comprehensive MoD-DNN framework.

Notations: Lower (upper)-case bold characters are used to denote vectors (matrices), and the vectors are by default in column orientation. $\mathrm{E}[\cdot]$ denotes the expected value of a discrete random variable. The superscripts $(\cdot)^{\mathsf{T}}$, and $(\cdot)^{\mathsf{H}}$ represent the transpose and conjugate transpose operators, respectively. Symbols $\odot$ stands for Hadamard product. $\jmath = \sqrt{-1}$ is the imaginary unit.

Related Work

Recently, the success of machine learning across a wide range of disciplines gave rise to NN-aided AoA estimation. With the continuous enhancement of computational power, NN have widespread application in localization systems.

End-to-End Learning Numerous studies utilize deep CNNs to directly map input signals to estimated AoA parameters in an end-to-end fashion (Xuan et al. 2023). Augmenting the DNN architecture enables both offline and online learning for AoA estimation (Huang et al. 2018), with multiple hidden layers enhancing recognition potential and sparse feature extraction within the angular domain (Gehring et al. 2021). However, their efficacy hinges on NN's generalization capacity (Kotary et al. 2021), necessitating substantial training data and associated labels.

Feature Learning Alternative methodologies incorporate a feature learning strategy (Hou et al. 2019), yielding a subtle improvement in physical interpretability. By leveraging neural networks, specific input parameter features (Naseri et al. 2022) are acquired and refined, subsequently applied for AoA estimation, thereby reinforcing the linkage between input and estimated parameters with discernible physical significance. Nonetheless, these approaches, centered solely on training a single parameter (Jiang et al. 2019), run the risk of omitting valuable information pertinent to other critical feature parameters.

Model-Driven Learning Diverging from pure data-driven neural networks, certain methodologies integrate principled mathematical models with data-driven systems, reaping the strengths of both paradigms (Zhou, Li, and Wang 2020; Zhu et al. 2020; Fan et al. 2023). Hybrid model-driven deep learning schemes have been advanced to synergize prior statistical models (Merkofer et al. 2022; Hasanzade-Zonuzy, Kalathil, and Shakkottai 2021). These hybrid methods strategically harness partial domain knowledge provided by mathematical structures tailored to specific problems, while also leveraging learning from limited data.

Problem Formulation

Ideal Signal Model To begin with, we consider the ideal signal output in an uplink positioning scenario. Without loss of generality, the field of view is quantified using uniform discrete AoA set $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_L]^{\mathsf{T}}$ with a size of L, the antenna array of 5G gNB is an M-element uniform linear array (ULA) with unit inter-element spacing equivalent to half wavelength. Assuming that the uniform frequency increment of the orthogonal frequency division multiplexing (OFDM) signal with K-subcarrier is much smaller than the

central carrier frequency, the signal of the k-th subcarrier is expressed as

$$\begin{aligned} \mathbf{x}_k(t) &= \sum\nolimits_{l=1}^{L} \mathbf{a}(\theta_l) s_{l,k}(t) + \mathbf{n}_k(t) \\ &= \mathbf{A}(\boldsymbol{\theta}) \mathbf{s}_k(t) + \mathbf{n}_k(t), \quad t = 1, \cdots, T, \end{aligned} \quad (1)$$

where $s_{l,k}(t)$ represents the component on the k-th subcarrier of the sounding reference signal (SRS) transmitted by the UE locating at azimuth θ_l. $\mathbf{n}_k(t)$ denotes the additive white Gaussian noise (AWGN). $\mathbf{a}(\theta_l) = [1, \cdots, \mathrm{e}^{\jmath\pi(M-1)\sin\theta_l}]^{\mathsf{T}}$ and $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \cdots, \mathbf{a}(\theta_L)]$ are the steering vector and array manifold, respectively. Note that, the above ideal array manifold $\mathbf{A}$ is exactly determined by the azimuth of the UE and the ideal array configuration. By implementing fast Fourier transform (FFT), the channel state information (CSI) of the k-th subcarrier is derived as

$$\mathbf{h}(k) = \mathbf{A}(\boldsymbol{\theta}) \breve{\mathbf{s}}(k) + \breve{\mathbf{n}}(k), \quad k = 1, \cdots, K, \quad (2)$$

where $\breve{\mathbf{s}}(k)$ and $\breve{\mathbf{n}}(k)$ are signal and noise in frequency domain. In this case, $\mathbf{h}(k)$ is equivalent to a single snapshot signal. As hardware impairments generally exist in real-world systems, leading to the mismatch of the steering vectors $\tilde{\mathbf{a}}(\theta)$ in the real system and the ideal $\mathbf{a}(\theta)$, the performance of purely model-driven AoA estimators based on the ideal signal model usually degrade in practice.

Effect of Hardware Impairments Several types of error are proposed to model the effect of hardware impairments, which are the element position error, gain and phase inconsistencies, and mutual coupling (Pan et al. 2023). In most cases, the hardware impairments lead to phase error versus different AoAs. To precisely model the phase error into the signal model, the CSI-form receive signal of a practical system can be expressed as

$$\begin{aligned} \mathbf{h}(k) &= \sum\nolimits_{l=1}^{L} \boldsymbol{\gamma}(\theta_l) \odot \mathbf{a}(\theta_l) \breve{s}_l(k) + \breve{\mathbf{n}}(k) \\ &= \boldsymbol{\Gamma} \odot \mathbf{A}(\boldsymbol{\theta}) \breve{\mathbf{s}}(k) + \breve{\mathbf{n}}(k), \end{aligned} \quad (3)$$

where $\boldsymbol{\gamma}(\theta_l) = [\gamma_1(\theta_l), \cdots, \gamma_M(\theta_l)]^{\mathsf{T}}$ and error matrix $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}(\theta_1); \cdots; \boldsymbol{\gamma}(\theta_L)]$. The coefficient $\gamma_m(\theta_l)$ denotes the phase error of the m-th antenna for azimuth θ_l. Thus, the practical steering vector $\tilde{\mathbf{a}}(\theta_l) = \boldsymbol{\gamma}(\theta_l) \odot \mathbf{a}(\theta_l)$, and practical array manifold $\tilde{\mathbf{A}}(\boldsymbol{\theta}) = [\tilde{\mathbf{a}}(\theta_1), \cdots, \tilde{\mathbf{a}}(\theta_L)]$. Formula (3) reveals that the hardware impairment results in angular-dependent and nonlinear phase error in practical system.

Network Framework

To overcome the aforementioned influence, we propose a MoD-DNN framework with the ability of spectrum calibration to enhance AoA estimation performance. A sketch of the proposed framework is illustrated in Figure 1. In the preprocessing module, we first reformulate the input signal to a co-array signal form. Then, the MoD-DNN reconstruct the sparse spatial spectrum via iterative optimization between 1D-CNN and SCG. In the following, we respectively introduce those modules of the proposed framework in detail.

Preprocessing

Coarray Spatial Spectrum Considering the discrete AoA set $\boldsymbol{\theta}$, the covariance matrix of $\mathbf{h}(k)$ in practical system is
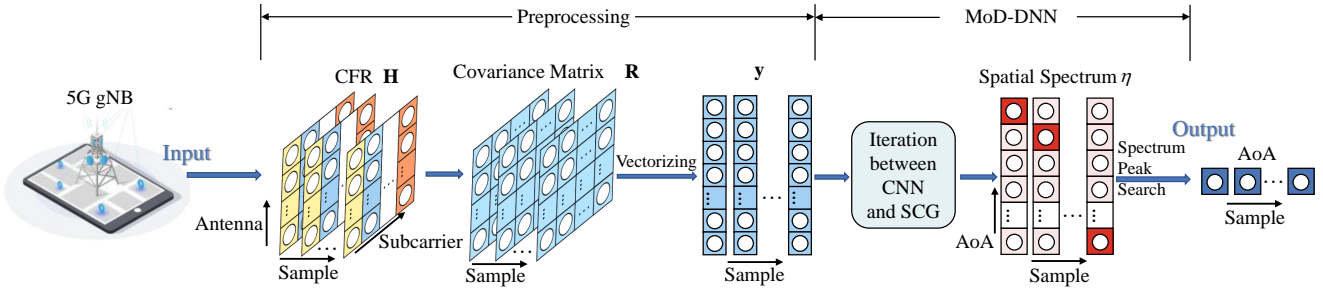
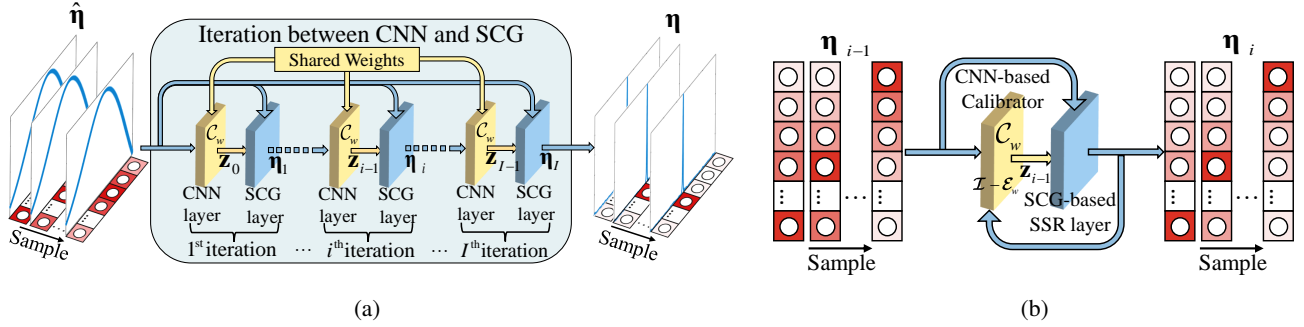Figure 1: Framework of MoD-DNN for AoA estimation.

(a)

(b)

Figure 2: Illustration of proposed iteration for MoD-DNN module. (a) Overall iteration in MoD-DNN module. (b) Alternative optimization between CNN-based calibrator and inverse problem-based SSR.

drived as

$$\mathbf{R} = \mathrm{E}\left[\mathbf{h}(k)\mathbf{h}^{\mathsf{H}}(k)\right] = \sum_{l=1}^{L} \eta_l \tilde{\mathbf{a}}(\theta_l)\tilde{\mathbf{a}}^{\mathsf{H}}(\theta_l) + \sigma_{\mathrm{n}}^2\mathbf{I}, \quad (4)$$

where η_l represents the signal power on θ_l, σ_n^2 is the power of the AWGN, and $\mathbf{I}$ denotes the identity matrix. Vectorizing the covariance matrix by stacking all columns of $\mathbf{R}$, we obtain the corresponding coarray signal $\mathbf{y} = \tilde{\mathbf{A}}\boldsymbol{\eta} + \sigma_{\mathrm{n}}^2\mathbf{e}$, where $\tilde{\mathbf{A}} = [\tilde{\mathbf{A}}_1; \cdots; \tilde{\mathbf{A}}_M]$ is the equivalent coarray manifold and $\tilde{\mathbf{A}}_m = [\tilde{\mathbf{a}}(\theta_1)\tilde{\mathbf{a}}^{\mathsf{H}}(\theta_1)\mathbf{e}_m, \cdots, \tilde{\mathbf{a}}(\theta_L)\tilde{\mathbf{a}}^{\mathsf{H}}(\theta_L)\mathbf{e}_m]$. Notice that, $\tilde{\mathbf{A}}_m$ is not equal to the ideal case $\mathbf{A}_m = [\mathbf{a}(\theta_1)\mathbf{a}^{\mathsf{H}}(\theta_1)\mathbf{e}_m, \cdots, \mathbf{a}(\theta_L)\mathbf{a}^{\mathsf{H}}(\theta_L)\mathbf{e}_m]$, and $\boldsymbol{\eta} = [\eta_1, \cdots, \eta_L]$ can be considered as the sparse spatial spectrum to be estimated. $\mathbf{e} = [\mathbf{e}_1; \cdots; \mathbf{e}_M]$ is a transition vector, where $\mathbf{e}_m \in \mathbb{R}^M$ is a vector with the m-th element being 1 whereas the others are zero. Generally, the CSS can be estimated using digital beamforming as

$$\hat{\boldsymbol{\eta}} = \mathbf{A}^{\mathsf{H}}\mathbf{y} \approx \mathbf{A}^{\mathsf{H}}\tilde{\mathbf{A}}\boldsymbol{\eta} = \tilde{\mathbf{P}}\boldsymbol{\eta}. \quad (5)$$

In this vain, the problem is equivalent to solve $\boldsymbol{\eta}$ from the observed spectrum $\hat{\boldsymbol{\eta}}$ with the projection matrix $\tilde{\mathbf{P}} = \mathbf{A}^{\mathsf{H}}\tilde{\mathbf{A}}$, and $\mathbf{A} = [\mathbf{A}_1; \cdots; \mathbf{A}_M]$.

Inverse Problem At this moment, we have reformulated the AoA estimation problem to a inverse problem. However, the projection matrix $\tilde{\mathbf{P}}$ usually differs from the ideal one $\mathbf{P} = \mathbf{A}^{\mathsf{H}}\mathbf{A}$ and is unknown with existence of error matrix $\boldsymbol{\Gamma}$ as per (3). In fact, the $\gamma_m(\theta_l)$ is difficult to be estimated and compensated for by traditional methods. To estimate the AoA from the observed coarray signal in practical system, we propose a model-driven network with iterations between deep learning and model-driven signal processing methods.

MoD-DNN Module

Iterative Optimization As shown in the overall iteration illustrated in Figure 2(a), the AoA resolution of the CSS is relatively low, and the position of the peak is inaccurate. To enhance the AoA estimation performance, the reconstruction of the spatial spectrum is expressed as the following optimization problem

$$\boldsymbol{\eta} = \arg\min_{\boldsymbol{\eta}} \|\mathbf{P}\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}\|_2^2 + \lambda\|\mathcal{E}_{\mathbf{w}}(\boldsymbol{\eta})\|^2, \quad (6)$$

where $\mathcal{E}_{\mathbf{w}}$ is a regularization term denoting the combination of noise and hardware impairments, which can be learned by training a $\mathbf{w}$-weighted CNN. The regularization coefficient λ is set as a trainable parameter as trade-off between calibration and reconstruction and can be fixed during the iterations. As such, the regularization term can be further expressed as $\mathcal{E}_{\mathbf{w}}(\boldsymbol{\eta}) = (\mathbf{I} - \mathcal{C}_{\mathbf{w}})(\boldsymbol{\eta}) = \boldsymbol{\eta} - \mathcal{C}_{\mathbf{w}}(\boldsymbol{\eta})$, where $\mathcal{C}_{\mathbf{w}}(\boldsymbol{\eta})$ is a calibrated version of $\boldsymbol{\eta}$ after the repairment of hardware impairments and noise. By utilizing the method of alternative optimization, the problem (6) is equivalent to the following iterations

$$\mathbf{z}^i = \mathcal{C}_{\mathbf{w}}(\boldsymbol{\eta}^i), \quad (7a)$$

$$\boldsymbol{\eta}^{i+1} = \arg\min_{\boldsymbol{\eta}} \|\mathbf{P}\boldsymbol{\eta} - \boldsymbol{\eta}^i\|_2^2 + \lambda\|\boldsymbol{\eta} - \mathbf{z}^i\|^2. \quad (7b)$$

The illustration of the iteration is depicted in Figure 2(b).

SCG Algorithm Now we introduce the SCG algorithm used in the MoD-DNN module. We reconstruct the spatial spectrum $\boldsymbol{\eta}$ by solving the sub-problem (7b) using a forward model. Conjugate gradient (CG) algorithm is usually utilized to solve such problem efficiently. All CG-steps have closed-form solutions which means no parameter needs extra training, and the gradients can be backpropagated from

Algorithm 1: SCG-based SSR algorithm.

Input: Coarray spectrum $\boldsymbol{\eta}^i$, calibrated spectrum $\mathbf{z}^i$, projection matrix $\mathbf{P}$, approximation parameter ϵ.
Output: Reconstructed spatial spectrum $\boldsymbol{\eta}^{i+1}$.
Initialize: $\boldsymbol{\eta}^{i+1}(0) = \mathbf{0}$, $\mathbf{g}(0) = (\mathbf{P} + \lambda\mathbf{I})\boldsymbol{\eta}^{i+1}(0) - (\boldsymbol{\eta}^i + \lambda\mathbf{z}^i)$, $\mathbf{c}(0) = -\mathbf{g}(0)$, γ_{CG} and $N_{\max}^{\mathrm{CGiter}}$.

1: **for** $n = 0$ to $N_{\max}^{\mathrm{iter}}$ **do**
2: $\quad \alpha(n) = -\frac{\mathbf{g}^{\mathrm{T}}(n)\mathbf{c}(n)}{\mathbf{c}^{\mathrm{T}}(n)\mathbf{P}\mathbf{c}(n)}$
3: $\quad \boldsymbol{\eta}^{i+1}(n+1) = \boldsymbol{\eta}^{i+1}(n) + \alpha(n)\mathbf{c}(n) - \nabla^s s(\boldsymbol{\eta}^{i+1}(n))$
4: $\quad \mathbf{g}(n+1) = (\mathbf{P} + \lambda\mathbf{I})\boldsymbol{\eta}^{i+1}(n) - (\boldsymbol{\eta}^i + \lambda\mathbf{z}^i)$
5: $\quad \beta(n) = \frac{(\mathbf{g}(n+1)-\mathbf{g}(n))^{\mathrm{T}}\mathbf{g}(n+1)}{\mathbf{g}^{\mathrm{T}}(n)\mathbf{g}(n)}$
6: $\quad \mathbf{c}(n+1) = -\mathbf{g}(n+1) + \beta(n)\mathbf{c}(n)$
7: $\quad\quad$ **if** $\boldsymbol{\eta}^{i+1}(n+1) - \boldsymbol{\eta}^{i+1}(n) < \gamma_{\mathrm{CG}}$ **then**
8: $\quad\quad\quad$ Break;
9: $\quad\quad$ **end if**
10: **end for**

---

the CG sub-blocks in iterations. By integrating the CG algorithm, the MoD-DNN module has sub-blocks consisting of numerical optimization layer with low cost of training.

Although the CG algorithm can efficiently solve the problem, we recall that the original form of the inverse problem is $\mathbf{y} = \tilde{\mathbf{A}}\boldsymbol{\eta}$, where $\tilde{\mathbf{A}}$ is an over-complete coarray manifold corresponding to the discrete AoA set $\boldsymbol{\theta}$. This implies that the inverse problem is underdetermined which leads to wide beam-width for the solution using CG algorithm. To further enhance the reconstruction of CG method, we propose a sparsity-constrained CG algorithm for spectrum reconstruction. In order to fully utilize the sparsity of $\boldsymbol{\eta}$, we further regularized the optimization problem (7b) with a sparsity function $s(\boldsymbol{\eta})$ as

$$\boldsymbol{\eta}^{i+1} = \arg\min_{\boldsymbol{\eta}} \|\mathbf{P}\boldsymbol{\eta} - \boldsymbol{\eta}^i\|_2^2 + \lambda\|\boldsymbol{\eta} - \mathbf{z}^i\|^2 + \mu \cdot s(\boldsymbol{\eta}), \quad (8)$$

where μ is the regularization coefficient with respect to sparsity constraint. Then, the minimization problem can be solved in an iterative method between CG solutions and sparsity modification. Without loss of generality, we can use reweighted zero attracting function as

$$s(\boldsymbol{\eta}(n)) = \log\left(1 + \|\boldsymbol{\eta}(n)\|_1/\epsilon\right), \quad (9)$$

where ϵ is an approximation parameter. The corresponding subgradient function in expressed as

$$\nabla^s s(\boldsymbol{\eta}(n)) = \frac{\mathrm{sgn}(\boldsymbol{\eta}(n))}{1 + \epsilon\|\boldsymbol{\eta}(n)\|_1}, \quad (10)$$

where $\mathrm{sgn}(\cdot)$ is the signum function. We summarize the SCG-based spatial spectrum reconstruction (SSR) algorithm in Algorithm 1.

Training of CNN-based Calibrator In this subsection, we introduce the 1D-CNN structure which is used as the spectrum calibrator in (7a) to effectively calibrate non-linear hardware impairments widespread in real-world systems. Unlike the mapping relationships between array output and angle information in end-to-end learning method (Yang et al.
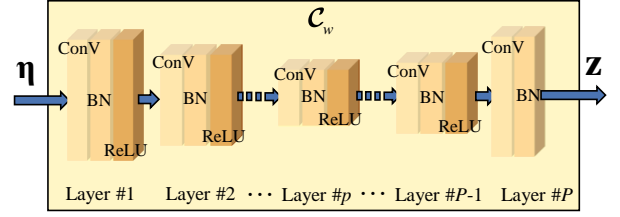

Figure 3: Structure of CNN-based calibrator.

2021), our approach transforms CSI (manifold format data) into image-format spatial spectrum. This conversion makes it well-suited for processing through CNNs. The structure of the 1D-CNN is given in Figure 3. Considering a training data set $\mathbb{D} = \{(\mathbf{y}(1), \boldsymbol{\eta}(1)), \cdots, (\mathbf{y}(D), \boldsymbol{\eta}(D))\}$. The coarray signal $\mathbf{y}(d)$ can be obtained from the CSI-form receive signal using (4), and the CSS $\hat{\boldsymbol{\eta}}(d)$ is generated using (5) and input to the CNN. Then, P layers are used to calibrate the spatial spectrum. In this experiment, 4 convolutional layers with convolutional kernel size of 32×1 are used to extract data features, and the number of kernels in each layer is 4, 8, 4, and 1, respectively. In addition, except for the activation function of the fourth layer, which uses Linear, the activation function of other layers uses ReLU. We select these hyperparameters through ablation experiments using a controlled variable methodology. The input of CNN in the i-th iteration is $\boldsymbol{\eta}^{i-1}$, and the calibrated spectrum is $\mathbf{z}^{i-1}$. Finally, $\mathbf{z}^{i-1}$ is input to the SCG algorithm to solve the sparse solution $\boldsymbol{\eta}^i$. When the iteration between the CNN and SCG modules is finished, the reconstructed spectrum $\boldsymbol{\eta}^I$ is the output of the framework. The goal of network training is minimizing the mean square error (MSE) loss function as

$$\mathcal{L} = \sum_{d=1}^{D} \|\boldsymbol{\eta}^I(d) - \boldsymbol{\eta}(d)\|^2. \quad (11)$$

For the network training in PyTorch, the Adam optimizer and StepLR scheduler are utilized to optimize the MSE loss, which allows the computation of the gradients of the weights using backpropagation.

Experimental Results

In this section, we demonstrate the effectiveness of AoA estimation with existence of hardware impairments using the proposed MoD-DNN, especially in comparison with multiple signal classification (MUSIC) algorithm, DeepMUSIC algorithm (Elbir 2020) and CNN (Liu et al. 2023). All novel datasets and codes introduced in this paper will be made publicly available upon publication of the paper with a license that allows free usage for research purposes.

Numerical Simulations

We consider a simulation scenario where a 5G gNB receives SRS signals transmitted by a user equipment (UE) in LOS condition. The CSI data is generated by the latest link-level 5G communications simulator (Jia et al. 2023). The simulator can incorporate customized hardware impairment functions and simulate channels in line with the 3GPP TR 38.901 standard. The indoor factory LoS channel at sub-
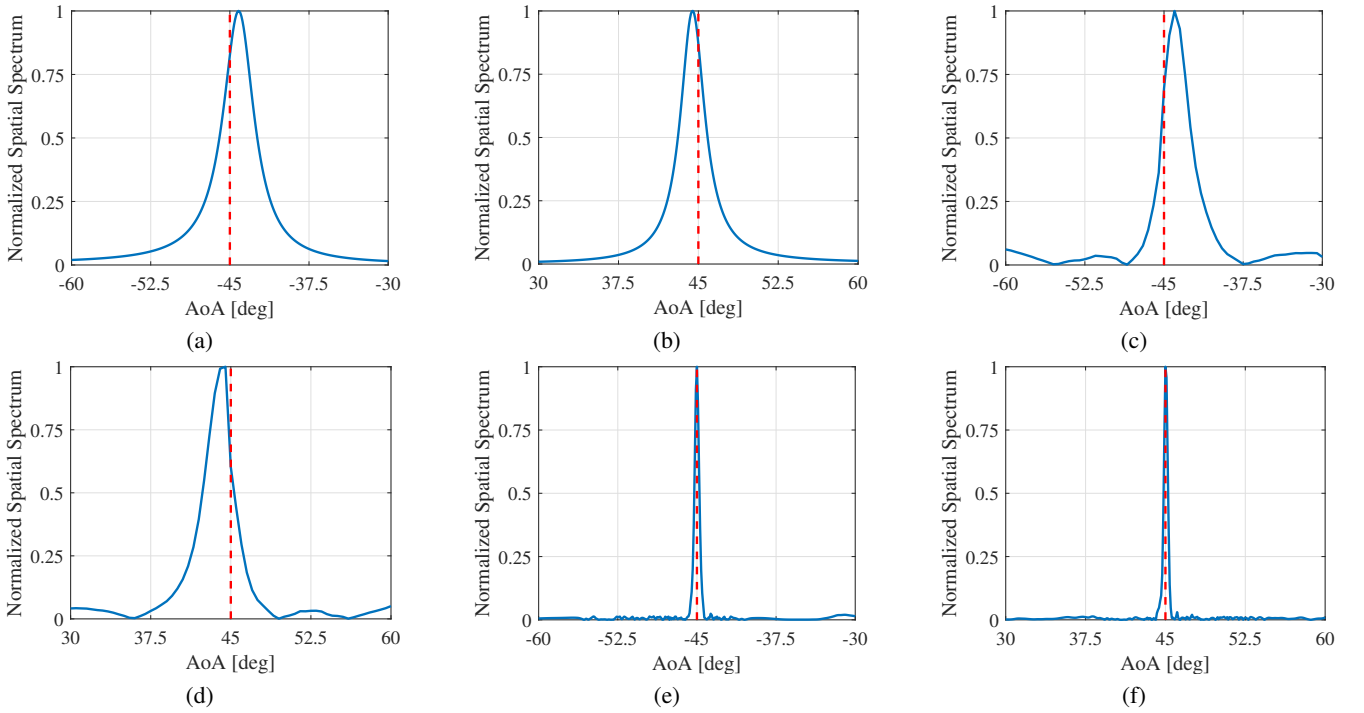
Figure 4: Spatial spectrum results at different AoA. Red dashed lines denote the truth. (a) MUSIC, $-45°$. (b) MUSIC, $45°$. (c) DeepMUSIC, $-45°$. (d) DeepMUSIC, $45°$. (e) MoD-DNN, $-45°$. (f) MoD-DNN, $45°$.

6GHz working frequency (3GPP 38.901 InF LoS) is chosen throughout the simulations.

Simulation Settings The central carrier frequency of the OFDM signal is f_c, and uniform frequency increment is Δf. The distance of UE from 5G gNB is fixed at R. The AoA of UE varies in θ with angular interval $\Delta\theta$. The parameter settings are listed in Table 1. Note that, B and P_{UE} respectively denote system bandwidth and power of UE. For each AoA, 45 SRS symbols are generated with the existence of error matrix Γ by the simulator. As such, We extract $1201 \times 40 = 48040$ groups of data for network training and $1201 \times 5 = 6005$ for validation.

Spatial Spectrum Results In the first set of simulations, the spatial spectrum results of MUSIC, DeepMUSIC and MoD-DNN at $-45°$, $45°$ are presented in Figure 4, in which signal-to-noise ratio (SNR) is fixed to 10 dB. It can be observed that MUSIC, DeepMUSIC and MoD-DNN can yield the spatial spectrum with a single peak which indicates the AoA of UE. However, due to the spectrum, the AoAs estimated by MUSIC and DeepMUSIC obviously deviate

Symbol	Value	Symbol	Value
f_c	4.8498 GHz	R	10 m
Δf	60 KHz	P_{UE}	23 dBm
B	100 MHz	$\Delta\theta$	0.1°
M	4	θ	$[-60°, 60°]$

Table 1: Simulation settings.

from the true value. As one of model-driven deep learning schemes, DeepMUSIC's performance is constrained without appropriate calibration for phase errors. Meanwhile, the spatial spectrum of the DeepMUSIC algorithm exhibits more fluctuations compared to the original MUSIC algorithm, which indicates a more severe impact of phase errors. As a comparison, the proposed MoD-DNN has the sharpest peak which accurately estimate the AoA.

RMSE of AoA Estimation In order to further compare the estimation accuracy, we observe the RMSE of different algorithms versus SNR. As shown in Figure 5(a), the performance of MUSIC and DeepMUSIC almost remain unchanged as SNR varies. This phenomenon indicates that these algorithm are statistically invalid with the existence of hardware impairments. Different from that, the RMSEs of CNN and the proposed MoD-DNN both decrease while MoD-DNN consistently outperforms the CNN. Compared to the purely data-driven CNN, the proposed MoD-DNN can effectively utilize the knowledge of signal model. Accordingly, we observe that the advantage of MoD-DNN becomes more obvious in high SNR region. To validate the effectiveness for impairment calibration of the proposed method, we further compare the RMSEs for different cases of impairment. Specifically, we use a weight coefficient ρ for error matrix Γ varies from 0 to 1 to simulate the degree of impairment. As illustrated in Figure 5(b), DeepMUSIC and MUSIC perform well under low impairment condition. However, the performances become worse as ρ increases. On the contrary, the performances of CNN and MoD-DNN only fluctuates slightly. Owing to the integration of NN and SCG, MoD-DNN successfully mitigates the influence of hardware
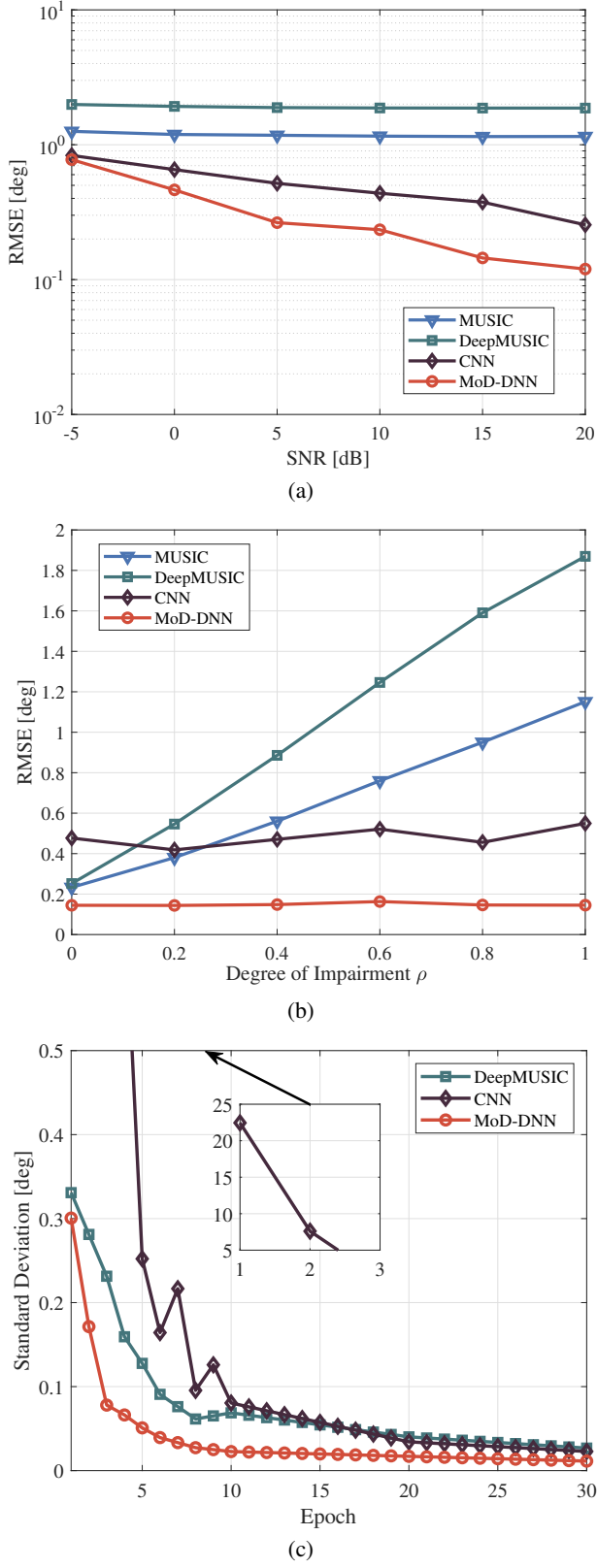
(a)

(b)

(c)

Figure 5: Performance comparison of different methods. (a) RMSE versus SNR. (b) RMSE versus degree of impairment ρ. (c) Standard deviation versus epoch.

impairments and outperforms the data-driven CNN.

Convergence Speed of Algorithms In the final simulation, we observe the standard deviation (SD) of loss versus epoch to evaluate the convergence speed of different algorithms. It is worth mentioning that we take the convergence value as the expectation to calculate the standard deviation. As shown in Figure 5(c), the SD of MoD-DNN converges rapidly compared to CNN and DeepMUSIC, which showcase the advantage of MoD-DNN with respect to training efficiency. So far, the effectiveness in terms of impairment calibration and estimation accuracy are validated by the numerical simulation.

Experiment in an Anechoic Chamber

In this subsection, we compare the AoA estimation performance of different methods in an anechoic chamber. Experiments conducted in an anechoic chamber provide valuable insights as NLoS and multipath propagation are absent. In this controlled environment, hardware impairments are the primary challenges to the accuracy of AoA estimation. The experimental setups are illustrated in Figure 6.
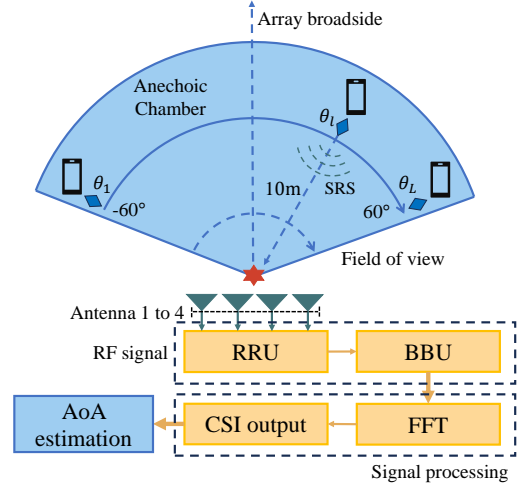

Figure 6: Setups for anechoic chamber experiments.

Experimental Settings We employs the UE and gNB for signal transmitting and receiving, respectively. The AoA of UE is rotated from $-60°$ to $60°$ by a uniform angle interval of $1°$ during the data acquisition process. For each AoA, 450 SRS symbols are transmitted by the UE. Then, the sounded CSIs are collected by the 5G gNB, yielding $121 \times 450 = 54450$ groups of CSI data. We extract $121 \times 400 = 48400$ groups of data for network training and $121 \times 50 = 6050$ for validation.

CDF of AoA Estimation Error The cumulative distribution function (CDF) curves of AoA estimation error are provided in Figure 7(a), which affirm the remarkable accuracy of the proposed MoD-DNN, outperforming all other methods. Particularly noteworthy is the MoD-DNN's capacity to achieve a minimum 95% reduction in the 80th percentile of AoA estimation error , from around $3°$ to $0.15°$.

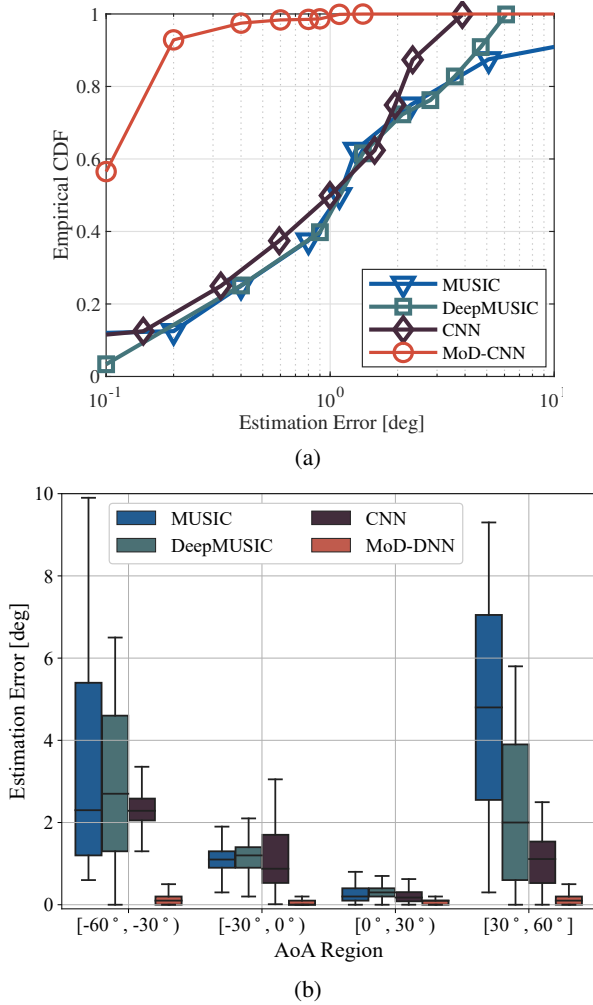**Boxplots of AoA Estimation Error** We proceed to evaluate the AoA estimation performance of various methods

Figure 7: AoA estimation error performance of different methods. (a) CDF. (b)Boxplots for 4 subregions.

through boxplots portraying estimation errors across varying angular sectors. Specifically, the field of view is segmented into four subregions: $[-60°, -30°), [-30°, -0°), [0°, 30°),$ and $[30°, 60°]$. The box's lower and upper boundaries correspond to the first quartile (Q1) and third quartile (Q3), respectively. The upper and lower whiskers respectively indicate the maximum and minimum estimation errors after removal of outliers (data with errors exceeding 1.5 times Q3). The median is represented by a horizontal line positioned above the box.

As displayed in Fig, 7(b), within the angular regions of $[-30°, -0°)$ and $[0°, 30°)$, the estimation results of all algorithms exhibit relatively accurate performance, with median errors of approximately less than 1°. This outcome is attributed to the minimal impact of angular-dependent phase errors in these specific regions. As the AoA of the User Equipment (UE) extends to the angular regions of $[-60°, -30°)$ and $[30°, 60°]$, the severity of phase errors escalates. Consequently, the MUSIC and DeepMUSIC algorithms experience significant performance degradation. For these two algorithms, the maximum errors are approximately 10° and 6.5°, respectively, while the medians hover

around 2°. The substantial fluctuation in phase errors also contributes to a high IQR for MUSIC and DeepMUSIC.In comparison, CNN and MoD-DNN are less affected by these conditions. Remarkably, the proposed MoD-DNN algorithm demonstrates superior performance across all assessed indicators represented by the box plots, encompassing maximum error, median error, and IQR in all angular regions. This phenomenon indicates a robust estimation performance of MoD-DNN and validates the effectiveness of error calibration for MoD-DNN.

Computational Complexity Lastly, the computational efficiency of MUSIC, DeepMUSIC, CNN, and MoD-DNN are compared. The running times are all recorded by the timer of PyTorch on a PC with Intel(R) Core(TM) i7-1065G7 CPU and a 16-GB RAM. The result is shown in TABLE. 2 while the training time is mentioned in hours, whereas the testing time is mentioned in milliseconds.

In comparison to algorithms generating spatial spectra as output, the CNN algorithm boasts a notably shorter training duration due to its direct output of low-dimensional AoA values. In the case of the MoD-DNN algorithm and the DeepMUSIC algorithm, the former capitalizes on the weight-sharing strategy within the CNN module and the closed-form solutions of the SCG algorithm. Consequently, the training parameters number of MoD-DNN is fewer than that of DeepMUSIC, resulting in a clear advantage in training time. In terms of testing time, it's clear that the computational time for all algorithms stays within the millisecond range, thus ensuring efficient real-time implementation.

	MUSIC	DeepMUSIC	CNN	**Ours**
training [h]	-	30.8	1.6	**15.6**
testing [ms]	9.4	31.6	1.8	**14.9**

Table 2: Training and testing times.

Conclusion

This study addresses AoA estimation in the presence of hardware impairments. We reformulate direction finding as spatial spectrum reconstruction through an inverse problem formulation. The iterative optimization between CNN and SCG enhances spatial spectrum calibration, leading to improved estimation precision. Simulation and experimental findings confirm the superiority of our proposed framework over both pure model-driven and data-driven approaches. Our insights into hybrid AI-and-model-driven methodologies hold promise for mitigating radio-frequency hardware-related challenges and enhancing the overall system reliability and efficiency for the broader spectrum of wireless tasks.

Acknowledgements

References

Dai, Z.; He, Y.; Tran, V.; Trigoni, N.; and Markham, A. 2022. DeepAoANet: Learning angle of arrival from software defined radios with deep neural networks. *IEEE Access*, 10: 3164–3176.

Elbir, A. M. 2020. DeepMUSIC: Multiple signal classification via deep learning. *IEEE Sens. Lett.*, 4(4): 7001004.

Fan, W.; Liu, S.; Li, C.; and Huang, Y. 2023. Fast direct localization for millimeter wave MIMO systems via deep ADMM unfolding. *IEEE Wireless Commun. Lett.*, 12(4): 748–752.

Gehring, C.; Kawaguchi, K.; Huang, J.; and Kaelbling, L. 2021. Understanding end-to-end model-based reinforcement learning methods as implicit parameterization. In *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 703–714. Online.

Ghourchian, N.; Allegue-Martinez, M.; and Precup, D. 2017. Real-time indoor localization in smart homes using semi-supervised learning. In *Proc. 31st Annu. AAAI Conf. Artif. Intell. (AAAI)*, 4670–4677. San Francisco, CA, USA.

HasanzadeZonuzy, A.; Kalathil, D.; and Shakkottai, S. 2021. Model-based reinforcement learning for infinite-horizon discounted constrained Markov decision processes. In *Proc. 30th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2519–2525. Online.

Hou, Y.; Ma, Z.; Liu, C.; and Loy, C. C. 2019. Learning to steer by mimicking features from heterogeneous auxiliary networks. In *Proc. 33rd Annu. AAAI Conf. Artif. Intell. (AAAI)*, 8433–8440. Honolulu, HI, USA.

Huang, H.; Yang, J.; Huang, H.; Song, Y.; and Gui, G. 2018. Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system. *IEEE Trans. Veh. Technol.*, 67(9): 8549–8560.

Huang, M.; Meng, F.; Liu, S.; Zhang, C.; Huang, Y.; and Lu, Z. 2023. TDoA positioning with data-driven LoS inference in mmWave MIMO communications. In *Proc. ACM SenSys Workshop mmWave Sens. Syst. App.*, 1–7. Istanbul, Türkiye.

Jia, X.; Liu, P.; Liu, S.; Li, X.; and Qi, W. 2023. Link-level simulator for 5G localization. *IEEE Trans. Wireless Commun.*, 22(8): 5198–5213.

Jiang, B.; Wu, X.; Yu, K.; and Chen, H. 2019. Joint semi-supervised feature selection and classification through Bayesian approach. In *Proc. 33rd Annu. AAAI Conf. Artif. Intell. (AAAI)*, 3983–3990. Honolulu, HI, USA.

Kotary, J.; Fioretto, F.; Van Hentenryck, P.; and Wilder, B. 2021. End-to-end constrained optimization learning: A survey. In *Proc. 30th Int. Joint Conf. Artif. Intell. (IJCAI)*, 4475–4482. Online.

Lee, H.; Kim, Y.; Seol, S.; and Chung, J. 2023. Deep learning-based direction-of-arrival estimation for far-field sources under correlated near-field interferences. *ICT Express*, 9(4): 741–747.

Liu, S.; Wang, H.; Pan, M.; Liu, P.; Ma, Y.; and Huang, Y. 2023. 5G NR monostatic positioning with array impairments: Data-and-model-driven framework and experiment results. In *Proc. 29th ACM MobiCom Workshop on Integr. Sens. Commun. Syst.*, 1–6. Madrid, Spain.

Merkofer, J. P.; Revach, G.; Shlezinger, N.; and van Sloun, R. J. G. 2022. Deep augmented MUSIC algorithm for data-driven DoA estimation. In *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 3598–3602. Singapore.

Mo, S.; and Morgado, P. 2023. A unified audio-visual learning framework for localization, separation, and recognition. In *Proc. 40th Int. Conf. Mach. Learn. (ICML)*, 25006–25017. Honolulu, HI, USA.

Naseri, M.; Shahid, A.; Gordebeke, G.-J.; Lemey, S.; Boes, M.; Van De Velde, S.; and De Poorter, E. 2022. Machine learning-based angle of arrival estimation for ultra-wide band radios. *IEEE Commun. Lett.*, 26(6): 1273–1277.

Pai, M.; and Kumar, A. 2019. Distribution learning of a random spatial field with a location-unaware mobile sensor. In *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 12447–12455. Vancouver, BC, Canada.

Pan, M.; Liu, P.; Liu, S.; Qi, W.; Huang, Y.; You, X.; Jia, X.; and Li, X. 2022. Efficient joint DOA and TOA estimation for indoor positioning with 5G picocell base stations. *IEEE Trans. Instrum. Meas.*, 71: 8005219.

Pan, M.; Liu, S.; P. Liu, W. Q.; Huang, Y.; Z. Wang, Q. W.; and Gardill, M. 2023. In situ calibration of antenna arrays for positioning with 5G networks. *IEEE Trans. Microwave Theory Tech.*, 71(10): 4600–4613.

Sun, S.; Melamed, D.; and Kitani, K. 2021. IDOL: Inertial deep orientation-estimation and localization. In *Proc. 35th Annu. AAAI Conf. Artif. Intell. (AAAI)*, 6128–6137. Online.

Xuan, T.; Zhu, Y.; Borca-Tasciuc, G.; Liu, M. X.; Sun, Y.; Dean, C.; Morales, Y. C.; Shi, Z.; and Yu, D. 2023. End-to-end pipeline for trigger detection on hit and track graphs. In *Proc. 37th Annu. AAAI Conf. Artif. Intell. (AAAI)*, 15752–15758. Washington, DC, USA.

Yang, M.; Ai, B.; He, R.; Huang, C.; Ma, Z.; Zhong, Z.; Wang, J.; Pei, L.; Li, Y.; and Li, J. 2021. Machine-learning-based fast angle-of-arrival recognition for vehicular communications. *IEEE Trans. Veh. Technol.*, 33(1): 1592–1605.

You, X.; Huang, Y.; Liu, S.; and et al. 2023. Toward 6G TKμ extreme connectivity: Architecture, key technologies and experiments. *IEEE Wireless Commun.*, 30(3): 86–95.

Zhou, Q.; Li, H.; and Wang, J. 2020. Deep model-based reinforcement learning via estimated uncertainty and conservative policy optimization. 6941–6948. New York City, NY, USA.

Zhu, G.; Zhang, M.; Lee, H.; and Zhang, C. 2020. Bridging imagination and reality for model-based deep reinforcement learning. In *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, volume 33, 8993–9006. Online.

Zhu, Y.; Xiao, Z.; Fang, Y.; Fu, X.; Xiong, Z.; and Zha, Z. 2022. Efficient model-driven network for shadow removal. In *Proc. 36th Annu. AAAI Conf. Artif. Intell. (AAAI)*, 3635–3643. Online.