One-Shot Near-Field Localization with AI-Optimized Hybrid Beamformer Design

One-Shot Near-Field Localization with AI-Optimized Hybrid Beamformer Design

Mattia Fabiani*‡, Davide Dardari*‡, Antonio A. D'Amico†‡, and Luca Sanguinetti†‡
*Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi", University of Bologna, Bologna, Italy.
†Department of Information Engineering, University of Pisa, Pisa, Italy.
‡National Laboratory of Wireless Communications (WiLAB), CNIT, 40136 Bologna, Italy

***Abstract*—This paper introduces a learning-based approach to the near-field source localization problem adopting a hybrid analog-digital beamformer in an extremely large-scale multiple-input multiple-output (XL-MIMO) system. Hybrid analog-digital architectures gained significant attention in the literature due to the limited number of radio-frequency (RF) chains. However, the investigation of effective techniques tailored to partially connected hybrid beamformers for near-field localization is still missing in the literature. To this end, we leverage a Convolutional Neural Network (CNN)-based model to: (*i*) perform the analog beamformer design with proper training constraints; and (*ii*) estimate the single-user near-field position in a single snapshot. In the inference stage, the model is divided into two parts: the first accounts for the beamformer design, and the second acts as a localizing function. Simulation results demonstrate superior performance of the proposed method over existing solutions and robustness in multipath propagation conditions. In addition, our network is scalable and requires fewer RF chains than fully-connected architectures.**

***Index Terms*—Convolutional neural network, near-field, XL-MIMO, one-shot localization, sub-array, inter-connected, hybrid beamforming.**

I. Introduction

Near-field source localization has gained interest with the rise of millimeter wave (mmWave) and (sub-)terahertz (THz) frequencies, driven by the need for precise positioning in next-generation wireless systems. As the Rayleigh distance extends to hundreds of meters, the traditional far-field assumption becomes less accurate [1]. Instead, near-field propagation enables joint estimation of both angle and range, enhancing localization and tracking accuracy.

Near-field source localization can be achieved using fully digital architectures, where the number of radio-frequency (RF) chains equals the number of antennas [2], [3]. However, in extremely large-scale MIMO (XL-MIMO) systems, this approach is impractical due to high cost and energy consumption. To address this, research has focused on hybrid architectures, which reduce hardware complexity by using fewer RF chains than antennas. Several studies have investigated the challenges of designing hybrid MIMO systems [4]–[8]. The authors of [7], [8] addressed the problem of hybrid beamformer design via iterative algorithms, leveraging radio-frequency (RF) chain selection based on current communication needs. However, they do not address near-field localization, which is considered in [2], [9] with a focus on multi-snapshot localization. This approach is unsuitable for real-time applications due to rapid environmental changes and limited time for position estimation. To overcome this challenge, deep learning has emerged as a powerful tool for capturing complex, non-linear relationships. A single-snapshot localization method was proposed in [10], where a deep neural network (DNN) optimizes the analog beamformer and serves as a localization function. However, this solution requires extensive connections between receiving antennas and RF chains, posing practical implementation challenges. Additionally, since analog beamformers allow for an infinite range of configurations, the authors benchmarked their model using only a single specific setup.

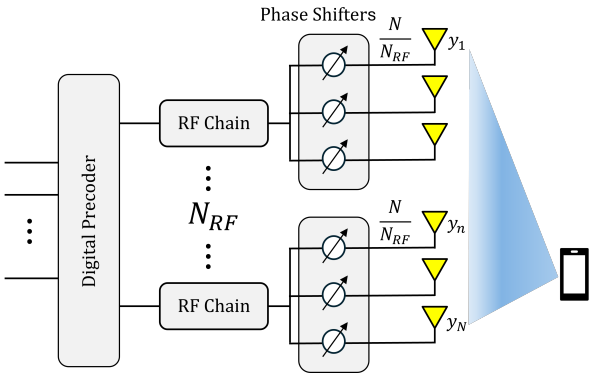
In this paper, we investigate various analog beamformer configurations, including sub-array and inter-connected architectures, where each RF chain is connected to a subset of antennas. Building on [10], we first design the analog beamformer through constrained training of fully connected layers. Then, we develop a localization function using a Convolutional Neural Network (CNN) followed by fully connected layers. To leverage the sparse nature of the mmWave channel, we extend our model to estimate multipath components from scatterers. The proposed method achieves superior localization accuracy compared to [10] while improving scalability by reducing the number of phase shifters in the analog beamformer.

All the simulation results can be reproduced using the Python code available at https://github.com/mattiafabiani/One-Shot-Near-Field-Localization-with-AI-Optimized-Hybrid-Beamformer-Design.

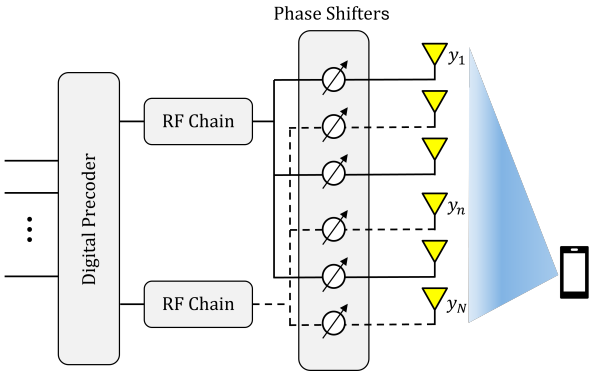
II. System Model

We consider the uplink of a narrowband XL-MIMO system. As shown in Fig. 1, the base station (BS) uses hybrid beamforming with M RF chains and a uniform linear array (ULA) with $N \geq M$ antennas, spaced $d = \frac{\lambda}{2}$ apart where λ denotes the wavelength. A single-antenna source is located at position $\mathbf{p} = [r\sin\theta, r\cos\theta]^T$, where r and θ are the range and angle of the source. Let ρ denote the received signal-to-

This work has been performed in the framework of the HORIZON-JU-SNS-2022 project TIMES, grant no. 101096307, co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union



(a) Sub-array



(b) Inter-connected

Fig. 1: Example of (a) sub-array and (b) inter-connected beamformer configurations, where $N = 6$, $M = 2$.

noise ratio (SNR) of the symbol transmitted by the source. Thus, the received signal at the BS is

$$\mathbf{y} = \sqrt{\rho}\,\mathbf{a}(\theta, r) + \mathbf{n} \tag{1}$$

where $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N) \in \mathbb{C}^{N\times 1}$ is thermal noise, $\mathbf{a}(\theta, r)$ denotes the array response vector given by [11]

$$\mathbf{a}(\theta, r) = [e^{-j\frac{2\pi}{\lambda}(r_1 - r)}, \ldots, e^{-j\frac{2\pi}{\lambda}(r_N - r)}]^T \tag{2}$$

with r_n being the distance between the source and antenna n, which can be computed as [11]:

$$r_n = \sqrt{r^2 + \delta_n^2 d^2 - 2r\sin\theta\delta_n d} \tag{3}$$

and $\delta_n = (2n - N - 1)/2$. The received signal $\mathbf{y}$ is then combined through the analog beamformer $\mathbf{V} \in \mathbb{C}^{M\times N}$ to obtain

$$\bar{\mathbf{y}} = \mathbf{V}\mathbf{y} = \sqrt{\rho}\mathbf{V}\mathbf{a}(\theta, r) + \mathbf{V}\mathbf{n} \tag{4}$$

where $|[\mathbf{V}]_{mn}| = \frac{1}{\sqrt{N}}$ to account for power balance.

Our goal is to design the analog combining matrix $\mathbf{V}$ to estimate the source position from the beamformed signal $\bar{\mathbf{y}}$, while reducing the NM phase shifters required in fully connected designs [10]. To achieve this, we explore two alternative hybrid structures: subarray [12] and inter-connected architectures. In this way, only N phase shifters are required.

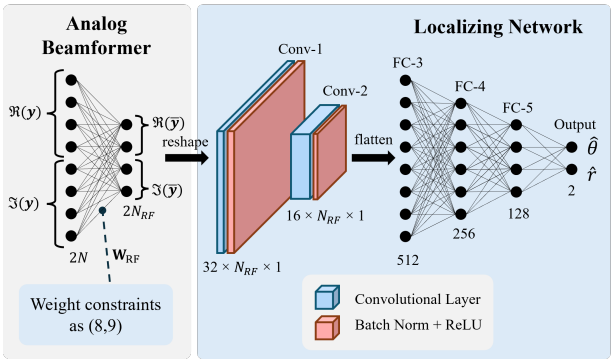


Fig. 2: Structure of the developed DNN, divided into two parts: analog beamformer and localizing function.

1) The first architecture is depicted in Fig. 1a, and it is such that each RF chain is connected to a subset of N/M adjacent antennas. This yields

$$\mathbf{V}^{(\text{sub})} = \text{blkdiag}(\mathbf{v}_1^T, \ldots, \mathbf{v}_M^T) \tag{5}$$

where $\mathbf{v}_i \in \mathbb{C}^{N/M}$ denotes the vector containing the weights of the analog phase shifters between the i^{th} RF chain and i^{th} subset of antenna elements.
2) The second architecture is an alternative setup where each RF chain is connected to N/M antennas with a uniform connection spacing of M. Specifically, the i^{th} RF chain is connected to the antennas indexed by $i + jM$, where $j = 0, \ldots, \frac{N}{M} - 1$.

III. Learning-Based Near-Field Localization

Building on [10], we explore the potential of DNNs to learn complex data-driven mappings. Here, the network is trained end-to-end to capture the relationship between the received signal at the ULA and the polar coordinates $(\hat{\theta}, \hat{r})$.

Since range and angle components serve as labels, a pre-processing step is required to ensure they contribute equally to the model's output. To achieve this, the polar coordinates of the near-field source are scaled as follows:

$$\begin{cases} r' = 2(r - r_{\min})/(r_{\max} - r_{\min}) - 1 \\ \theta' = \sin\theta \end{cases} \tag{6}$$

where $r_{\min}$ and $r_{\max}$ represent the minimum and maximum feasible range of the user. Both range and angle are bounded to $[-1, 1]$. The following subsections provide a detailed explanation of how the proposed DNN designs the analog beamformer and extracts localization information from the raw received signal. Finally, we present a complexity analysis of the proposed method.

A. Analog Beamformer Design

We being by describing the methodology to turn a fully-connected layer into a learnable analog beamformer. First, the complex-valued received signal $\mathbf{y} \in \mathbb{C}^{N\times 1}$ is split into the real and imaginary parts and then stacked into a real-valued column vector $\mathcal{F}\{\mathbf{y}\} = [\Re\{\mathbf{y}^T\}, \Im\{\mathbf{y}^T\}]^T \in \mathbb{R}^{2N\times 1}$.

As described in [10], the weights of a fully-connected layer cannot be employed directly to design a hybrid beamformer. To ensure the feasibility of the model, specific constraints must be applied during the training process. Let us denote $\bar{\mathbf{W}}^{(\text{in})} \in \mathbb{R}^{2N_{\text{RF}} \times 2N}$ as the weights of the first network layer. Then, $\mathbf{W}_1^{(\text{in})} \in \mathbb{R}^{N_{\text{RF}} \times N}$ and $\mathbf{W}_2^{(\text{in})} \in \mathbb{R}^{N_{\text{RF}} \times N}$ represent two quadrants of $\bar{\mathbf{W}}^{(\text{in})}$ and will serve as the real and imaginary part of the analog beamformer. Thus, the network is trained according to the following constraints

$$(\widetilde{\mathbf{W}}, \widetilde{\mathbf{b}}) = \arg\min_{\mathbf{W},\mathbf{b}} \mathbb{E}\left\{ \left\| \boldsymbol{\gamma}^{(i)} - f^{(\text{train})}(\mathbf{y}^{(i)}) \right\|_2^2 \right\} \quad (7)$$

subject to

$$f^{(\text{train})}(\mathbf{y}^{(i)}) = f^{(\text{out})}(\dots(f^{(1)}(\mathbf{b}^{(\text{in})} + \mathbf{W}^{(\text{in})}\mathcal{F}(\mathbf{y}^{(i)})))),$$

$$\mathbf{W}^{(\text{in})} = \begin{bmatrix} \mathbf{W}_1^{(\text{in})} - \mathbf{W}_2^{(\text{in})} \\ \mathbf{W}_2^{(\text{in})} \mathbf{W}_1^{(\text{in})} \end{bmatrix} \quad (8)$$

$$\left(\left[\mathbf{W}_1^{(\text{in})} \right]_{m,n} \right)^2 + \left(\left[\mathbf{W}_2^{(\text{in})} \right]_{m,n} \right)^2 = \frac{1}{N}, \quad (9)$$

$$\forall n \in \{1, \dots, N\}, \forall m \in \{1, \dots, M\}$$

where i indicates the index of a training sample, $\boldsymbol{\gamma}^{(i)} = [\theta'^{(i)}, r'^{(i)}]^T$ represents the ground-truth label, while $\mathbf{W} = \{\mathbf{W}^{(in)}, \mathbf{W}^{(1)}, \dots, \mathbf{W}^{(L)}\}$ and $\mathbf{b} = \{\mathbf{b}^{(\text{in})}, \mathbf{b}^{(1)}, \dots, \mathbf{b}^{(L)}\}$ denote the network weights and biases of the L hidden layers. After the training stage, the network weights and biases are represented as $\widetilde{\mathbf{W}}$ and $\widetilde{\mathbf{b}}$, respectively. The proof of the equality $\mathbf{W}^{(\text{in})}\mathcal{F}\{\mathbf{y}^{(i)}\} = \mathbf{W}_{\text{RF}}\mathbf{y}$ can be found in [10], and it allows the utilization of the trained weights to act as the analog phase shifters. The function $f^{(i)}(\mathbf{x})$ describes the forward pass and activation function at the i$^{\text{th}}$ layer, and it can be described as $f^{(i)}(\mathbf{x}) = f_{\text{act}}^{(i)}(\mathbf{W}^{(i)}\mathbf{x} + \mathbf{b}^{(i)})$, where $f_{\text{act}}^{(i)}$ is the non-linear activation function.

B. Localizing Network

The purpose of the second part of the network, as shown in Fig. 2, is to predict the near-field components of the user given the beamformed signal $\overline{\mathbf{y}}$. The real and imaginary parts of the beamformed signal are stacked and reshaped to fit the first convolutional layer, with 32 filters of size 1×1. The second convolutional layer applies 16 filters with a filter size of 1×1. Each CNN layer is followed by a batch normalization layer and ReLU activation function. Then, the output of the convolutional layer is vectorized and fed to a set of fully-connected layers of size $(512, 256, 128)$. Depending on the nature of the problem, the network output is a multiple of the number of paths, i.e. $n_{\text{out}} = 2L$, with tanh activation function to ensure the output to be in the range $[-1, 1]$. Then, the network output is converted into the near-field coordinates $[\hat{\theta}, \hat{r}]^T$ by inverting the functions in (6). The estimated position in cartesian coordinates can be derived as $\hat{\mathbf{p}} = [\hat{r}\cos\hat{\theta}, \hat{r}\sin\hat{\theta}]^T$.

C. Complexity Analysis

We analyze the computational complexity of the proposed hybrid beamforming approach in terms of floating-point operations (FLOPs) and compare it with the fully digital scheme, which utilizes the maximum likelihood algorithm. The complexity of the maximum likelihood is given by $2(4N+1)N_\theta N_r$ FLOPs, where N_θ and N_r denote the number of angle and range *bins*, respectively. In particular with $N = 128$, $N_\theta = 180$ and $N_r = 100$, the maximum likelihood approach requires $\approx 18.5 \times 10^6$ FLOPs. On the other hand, regarding the proposed CNN-based framework's complexity is given by

$$8N + 2\sum_{i=1}^{N_{\text{conv}}} h_{\text{out}}^{(i)} w_{\text{out}}^{(i)} C_{\text{out}}^{(i)} C_{\text{in}}^{(i)} K_1^{(i)} K_2^{(i)} + 2\sum_{i=1}^{N_{fc}} \gamma_i \gamma_{i-1}, \quad (10)$$

where the first term accounts for the first hidden layer of the network to model the hybrid beamformer architecture, and the remaining terms describe the complexity of the convolutional and fully-connected layers. N_{conv} is the number of convolutional layers, $h_{\text{out}}^{(i)}$ and $w_{\text{out}}^{(i)}$ are the height and width of the i$^{\text{th}}$ convolutional layer, $C_{\text{out}}^{(i)}$ and $C_{\text{in}}^{(i)}$ denote the i$^{\text{th}}$ output and input number of channels, and $K_1^{(i)} \times K_2^{(i)}$ is the i$^{\text{th}}$ filter size. Overall, with $N = 128$ antennas, $M = 16$ RF chains, the proposed method requires $\approx 0.77 \times 10^6$ FLOPs. Additionally, the sub-array and inter-connected beamforming architectures reduce the number of phase shifters by a factor of M, resulting in fewer connections in the first hidden layer compared to a fully connected architecture [10].

IV. Simulation Results

In this section, the performance accuracy of the proposed learning-based near-field localization approach is assessed by means of extensive simulations. We start by defining the simulation setup, followed by the simulation results.

A. Simulation Setup

We consider a near-field scenario as in [10], where a BS operating in the mmWave band, with a carrier frequency of $f_c = 25$ GHz, is equipped with a ULA comprising $N = 128$ antennas, each spaced $d = \frac{\lambda}{2}$. If not otherwise specified, the number of RF chains is set to $M = 16$. We first construct a large dataset with 10^5 channel realizations for different values of SNR. The user lies in the strong near-field region of the BS, and it is uniformly distributed in a $10\,\text{m} \times 10\,\text{m}$ area restricted to a range $r_{\text{min}} = 1$ m and $r_{\text{max}} = 10$ m. The Rayleigh distance with the system parameters is $d_R = \frac{2d^2(N-1)^2}{\lambda} \approx 100$ m. It is worth noting that we evaluate our network in the strong near-field region, which is in the order of $d_R/10$. This allows us to fully exploit the potential of near-field localization. Then, the dataset is split into 80% and 20% for training and validation sets. The maximum likelihood algorithm for the fully-digital system is employed with a grid of $N_\theta = 180$ and $N_r = 100$. The proposed model is generated with the PyTorch framework and trained on a PC with a 4-core CPU. During training, we utilize the Adam optimizer for 50 epochs with a batch size of 256, while the learning rate is 0.001. The hyperparameters have been tuned through trial and error procedures.

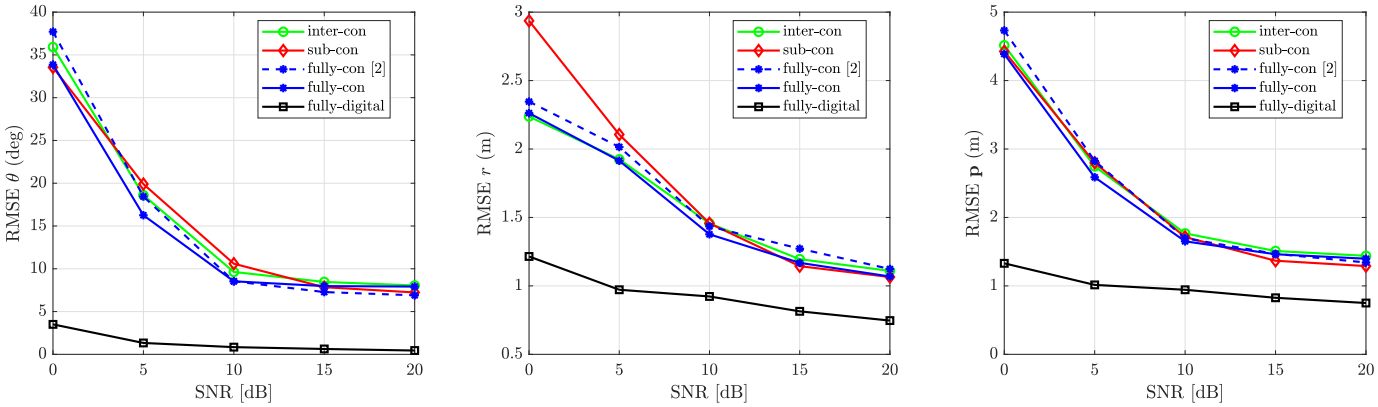


Fig. 3: RMSE of angle, range, and position versus SNR with different hybrid beamforming configurations, when $M = 16$ and $N = 128$.

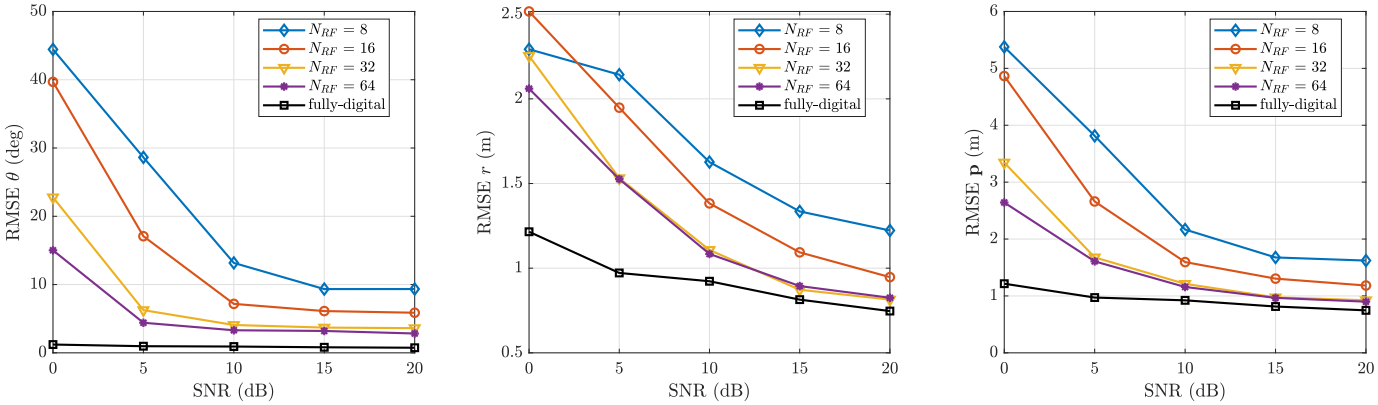


Fig. 4: RMSE of angle, range, and position versus SNR when the number of RF chains varies in a sub-array beamformer configuration with $M = 16$ and $N = 128$.

B. Numerical Results

This section shows the near-field localization performance of the proposed approach, evaluated with the root-mean-square error (RMSE) metric. The RMSE is computed for angle as $\sqrt{\mathbb{E}[(\theta - \hat{\theta})^2]}$, for range as $\sqrt{\mathbb{E}[(r - \hat{r})^2]}$, and for position as $\sqrt{\mathbb{E}[\|\mathbf{p} - \hat{\mathbf{p}}\|^2]}$. Since the predominance of the DNN-based approach has already been proved in [10], we focus on benchmarking the proposed solution with the fully-digital architecture.

1) Hybrid Architectures: Fig. 3 illustrates the angle, range, and position RMSE for diverse hybrid beamforming architectures, including a fully-digital configuration for benchmarking. The proposed CNN-based model demonstrates superior performance compared to the fully-connected beamformer structure presented in [10], achieving a relative improvement of 4.8%, which motivates the use of an improved localization network. The proposed sub-array and inter-connected structures perform similarly to the fully-connected architecture, while significantly reducing the number of phase shifters. Specifically, with the current configuration of $N = 128$ antennas and $M = 16$ RF chains, these methods reduce phase shifter requirements by a factor of 8. This reduction not only simplifies implementation but also addresses scalability challenges posed by the extensive interconnections in fully-connected designs [10].

2) Impact of Number of RF chains: It is noticeable in Fig. 3 a gap between the hybrid beamformers configurations with $M = 16$ RF chains and the fully-digital system. To this end, we explore the impact of the number of RF chains on localization accuracy. Fig. 4 shows the angle, distance, and position RMSE metric with $M = \{8, 16, 32, 64\}$ RF chains for a sub-connected analog beamformer configuration. As expected, the results show that as the number of RF chains increases, the overall RMSE decreases, reaching a positioning error as low as 0.9 m at 20 dB with 32 RF chains. However, a trade-off persists between the number of RF chains and the analog beamformer complexity.

C. Impact of Multipath

We now formulate the problem of near-field source localization in the presence of multipath propagation. At mmWave frequency bands, the severe path loss typically limits the number of significant multipath components. Therefore, we consider the strongest reflection component of the channel between the source and the BS, originating from a point-like scatterer in proximity to the user. The multipath channel with

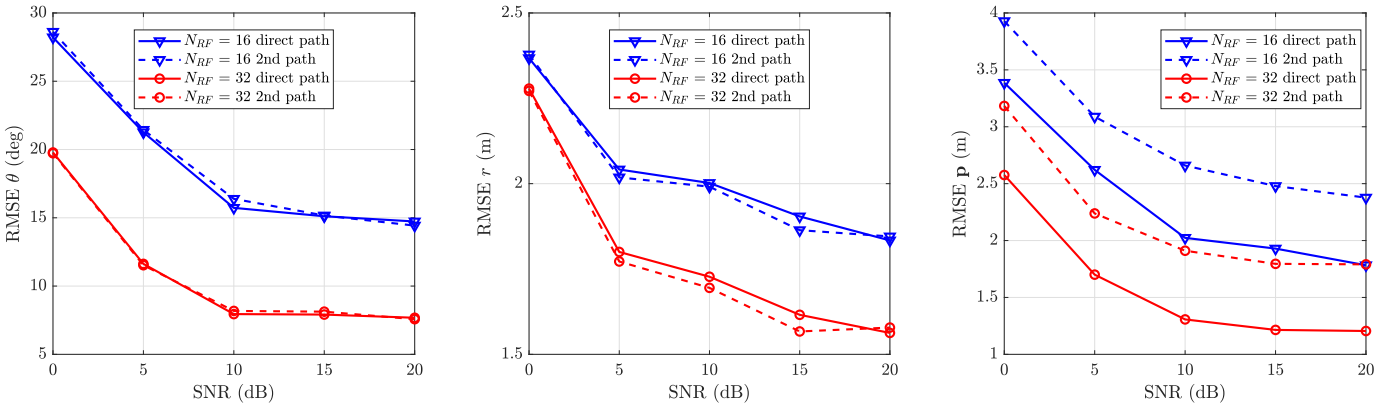


Fig. 5: RMSE of angle, range, and position versue SNR when the number of RF chains varies in a sub-array configuration with multipath.

P paths can be modeled as the summation of Line-of-Sight (LoS) and Non-Line-of-Sight (NLoS) components, which can be formulated as

$$\begin{cases} \mathbf{h}_{\text{LoS}} = \sqrt{\rho}\mathbf{a}(\theta, r) \\ \mathbf{h}_{\text{NLoS}} = \sum_{i=1}^{P} \sqrt{\rho_i} e^{j\phi_i} \mathbf{a}(\theta_i, r_i) \end{cases} \tag{11}$$

where θ_i and r_i denote the angle and distance between the i^{th} scatterer and the BS, while ρ_i and ϕ_i are the SNR and phase deviation of the i^{th} path. The scatterer is modeled such that it applies a random phase shift uniformly distributed in $\phi_i \in [-\pi, \pi]$ and a random amplitude response as $\mathcal{CN}(0, 1)$ due to the unknown geometry of the scatterer.

Similarly to (1), the received signal at the BS is

$$\mathbf{y} = \sqrt{\rho}\mathbf{a}(\theta, r) + \sqrt{\rho}_1 e^{j\phi_i} \mathbf{a}(\theta_1, r_1) + \mathbf{n} \tag{12}$$

where we consider a single scatterer, i.e., $P = 1$. The accuracy performance in the presence of multipath is illustrated in Fig. 5. We generate a point-like scatterer up to 1.5 meters from the source and aim to estimate the position of both the user and scatterer. To this end, we consider the hybrid interconnected structure with $M = \{16, 32\}$ and $N = 128$. The proposed approach demonstrates robustness against multipath components. Indeed, despite the stringent single-snapshot localization approach, we are able to locate both the near-filed user and roughly the position of the scatterer. It is worth noting that the lower scatterer's SNR, as modeled in (11), results in a higher position RMSE compared to the near-field user.

V. Conclusion

This paper presented a CNN-based framework for the one-shot near-field source localization problem in hybrid beamforming systems, where the analog beamformer design was performed through constrained training. The proposed network architecture outperforms existing learning-based methods in terms of localization accuracy while improving scalability by reducing the phase shifter requirements by a factor of M compared to fully-connected systems. The results demonstrate that sub-meter localization accuracy can be achieved in a single-snapshot manner, even with a limited number of RF chains. The proposed method also shows robustness in the presence of multipath components. Additionally, depending on specific network requirements, both the number of RF chains and the pilot length can be adjusted to further enhance localization accuracy.

References

[1] Y. Liu, Z. Wang, J. Xu, C. Ouyang, X. Mu, and R. Schober, "Near-field communications: A tutorial review," *IEEE Open Journal of the Communications Society*, 2023.

[2] W. Liu, J. Xin, W. Zuo, J. Li, N. Zheng, and A. Sano, "Deep learning based localization of near-field sources with exact spherical wavefront model," in *Proc. of European Signal Processing Conf. (EUSIPCO)*. IEEE, Sep. 2019, pp. 1–5.

[3] S. Jang, C. Im, H. Lee, and C. Lee, "A single-snapshot localization for monostatic fda-mimo radar," *IEEE Commun. Lett.*, vol. 26, no. 12, pp. 2899–2903, 2022.

[4] Y. Wang, X. Chen, Y. Cai, and L. Hanzo, "Ris-aided hybrid massive mimo systems relying on adaptive-resolution adcs: Robust beamforming design and resource allocation," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 3, pp. 3281–3286, 2021.

[5] L. Zhao, S. Wan, K. Kang, W. Feng, and H. Qian, "A general digital predistortion linearization scheme for hybrid beamforming system," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 1, pp. 654–663, 2022.

[6] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive mimo for hybrid precoding," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 3027–3032, 2019.

[7] Z. Zhang, Y. Liu, Z. Wang, J. Chen, and T. Q. Quek, "Dynamic MIMO architecture design for near-field communications," *IEEE Trans. Wireless Commun.*, 2024.

[8] A. Kaushik, J. Thompson, E. Vlachos, C. Tsinos, and S. Chatzinotas, "Dynamic RF chain selection for energy efficient and low complexity hybrid beamforming in millimeter wave MIMO systems," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 4, pp. 886–900, 2019.

[9] J. He, L. Li, T. Shu, and T.-K. Truong, "Mixed near-field and far-field source localization based on exact spatial propagation geometry," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3540–3551, 2021.

[10] S. Jang and C. Lee, "DNN-driven single-snapshot near-field localization for hybrid beamforming systems," *IEEE Trans. Veh. Technol.*, Feb. 2024.

[11] M. Cui and L. Dai, "Channel estimation for extremely large-scale MIMO: Far-field or near-field?" *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2663–2677, 2022.

[12] Z. Lu, Y. Han, S. Jin, M. Matthaiou, and T. Q. Quek, "Near-field channel reconstruction and user localization for elaa systems," in *2022 Int. Symp. on Wireless Commun. Syst. (ISWCS)*. IEEE, Oct. 2022, pp. 1–6.