Low-Latency Deep Learning Based Doppler Velocity Estimation for Integrated Sensing and Communication

Samarth Sharma Bhardwaj†, Aakanksha Tewari*, Syed Asrar Ul Haq*, Sumit J Darak*
*Indraprastha Institute of Information Technology Delhi, New Delhi, India
†Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016 India
E-mail: samarthsb23@iitk.ac.in, {aakankshat,syedh,sumit}@iiitd.ac.in

***Abstract*—In millimeter-wave Integrated sensing and communication (ISAC) systems, a high-resolution Doppler velocity estimation is crucial for localizing multiple tightly coupled mobile users in the environment for subsequent communication. Conventional subspace methods, such as the Estimation of Signal Parameters via Rotational Invariance Technique (ESPRIT), provide high resolution but incur high complexity and latency, limiting their use in real-time hardware. This work proposes a low-complexity end-to-end multi-layer perceptron (MLP) for Doppler estimation for IEEE 802.11ad ISAC waveform. The proposed MLP achieves up to 59.8% lower RMSE than classical ESPRIT while using fewer slow-time packets and up to 20.4× lower latency on the system-on-chip (SoC). A multi-model extension further improves robustness across signal-to-noise ratio (SNR), providing an additional 56% gain at low SNR by adaptively selecting model parameters. These results demonstrate that MLP-based estimation provides a scalable, efficient alternative to classical subspace methods for real-time ISAC systems.**

***Index Terms*—Deep Learning, Multi-layer perceptron (MLP), radar signal processing, ESPRIT, Zynq multi-processor system-on-chip**

I. Introduction

Millimeter-wave (mmW) Integrated Sensing and Communication (ISAC) systems employ radar signal processing (RSP) at the base station (BS) to sense the environment using the same waveform, spectrum and hardware intended for high-gain directional communication between BS and mobile users (MUs) [1–3]. In sensing, accurate and fast Doppler velocity estimation is critical for identifying static clutters and resolving multiple tightly spaced MUs for reliable communication via beamforming. While the conventional fast Fourier transform (FFT) offers Doppler estimation with low hardware complexity, it is only suitable for coarse estimation and cannot distinguish multiple closely-spaced targets [4, 5].The classical subspace methods, such as the Estimation of Signal Parameters via Rotational Invariance Technique (ESPRIT) and Multiple Signal Classification (MUSIC), are commonly used for Doppler velocity estimation as they offer good resolution at high signal-to-noise (SNR). However, they are computationally intensive and suffer from performance deterioration at low SNR. In such scenarios, large number of packets are needed which increases latency and complexity [5, 6].

This work is supported by the funding received from Chip to Startup (C2S, project no.: EE-9/2/2021-R&D-E) project from Ministry of Electronics and Information Technology (MeiTy), Government of India.

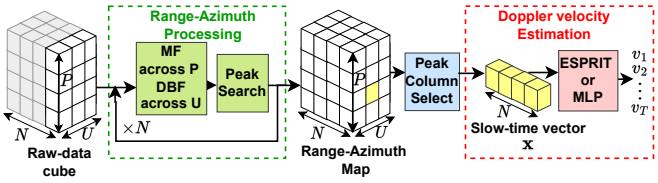

Fig. 1: RSP for ISAC system comprising of range, azimuth and Doppler estimation

Prior works have explored the integration of deep learning (DL) for enhancing the performance of RSP [7]. The works [7, 8] demonstrate that DL can effectively augment subspace methods for improved direction-of-arrival (DOA) estimation, offering better accuracy and resolution. DL-based approaches learn directly from data and do not rely on strict assumptions, such as additive white Gaussian noise (AWGN) or linear models, making them flexible for practical scenarios. However, their computational complexity and reliance on training data are concerns. In addition, the generalization capability of DL models remains a concern and must be validated across diverse operating environments, as well as their feasibility on SoCs.

We propose an end-to-end multi-layer perceptron (MLP) as a low-complexity solution replacing ESPRIT for Doppler velocity estimation in the ISAC RSP system, as shown in Fig. 1. The MLP model provides improved multiple-target localization compared to classical ESPRIT with fewer time packets. We present hardware-efficient architectures for MLP and ESPRIT on a Zynq multi-processor SoC (MPSoC) via hardware-software co-design (HSCD). The end-to-end MLP offers a 20.4× reduction in execution time over ESPRIT. The proposed multi-model MLP can dynamically adapt to SNR variations, achieving up to 56% improvement in performance by switching to the appropriate configuration.

II. ISAC System Model and RSP Framework

The ISAC system comprises a stationary BS and multiple MUs along with clutters as radar targets. We use IEEE 802.11ad as common waveform for sensing and communication [3, 4]. The BS omnidirectionally transmits preamble of the IEEE 802.11ad and processes the target-reflected returns using RSP to determine the range, azimuth, and Doppler velocity of the targets. The range-azimuth processing is carried out

through matched filtering (MF) across P fast-time samples and digital beamforming (DBF) across U antenna elements, as shown in Fig. 1 [4]. This is followed by Doppler estimation to separate out the MUs from static clutter scatterers. The slow-time vector, $\mathbf{x} \in \mathbb{C}^{N\times 1}$ post range-azimuth processing with superimposed returns from T targets within the same range-azimuth bin is represented as,

$$\mathbf{x}[n] = \sum_{t=1}^{T} a_t e^{-j\frac{4\pi}{\lambda}v_t n T_{PRI}} \tag{1}$$

where N is the number of slow-time samples, T_{PRI} is the pulse repetition interval (PRI), λ is the millimeter wavelength, a_t is the complex amplitude after reflection from the t^{th} target. $\mathbf{x}$ is processed through ESPRIT to determine the Doppler velocities of T targets. ESPRIT includes the autocovariance generation via spatial smoothing (SS), where $\mathbf{x}$ is split into multiple subarrays, each of length M, and the autocovariance resulting from each subarray is averaged to form the averaged covariance matrix, $\mathbf{A_{avg}} \in \mathbb{C}^{M\times M}$. $\mathbf{A_{avg}}$ then undergoes eigen vector decomposition (EVD) via the QR factorization method to determine the signal subspace, $\mathbf{E_s} \in \mathbb{C}^{M\times T}$, where T refers to the maximum number of targets we expect to detect in the signal. $\mathbf{E_s}$ is used for further processing via pseudo-inversion and eigenvalue decomposition to estimate the Doppler velocities of T sources.

EVD is the most computationally expensive step in subspace methods. This work and subsequent SoC implementation of ESPRIT use an iterative QR algorithm to enhance the accuracy of the EVD stage. In each k^{th} iterations, the covariance, $\mathbf{A_k} \in \mathbb{C}^{M\times M}$ is split into a unitary matrix $\mathbf{Q_k} \in \mathbb{C}^{M\times M}$ and an upper triangular matrix $\mathbf{R_k} \in \mathbb{C}^{M\times M}$. The covariance for the next iteration is reconstructed from the product $\mathbf{A_{k+1}} = \mathbf{R_k Q_k}$. These iterative QR decompositions are carried out until the magnitude of all of the off-diagonal elements of $\mathbf{A_k}$ falls below a threshold ϵ. The termination condition ensures that the resulting $\mathbf{A_{K-1}}$ converges to a diagonal matrix, with its diagonal elements now representing the eigenvalues of $\mathbf{A_{avg}}$. The respective eigenvectors are obtained by accumulating all the unitary matrices across iterations, $\mathbf{Q} = \mathbf{Q_0 Q_1 ... Q_{K-1}}$.

III. Proposed MLP Based Doppler Estimation

The performance of classical ESPRIT depends on signal quality and the number of slow time packets, and degrades significantly under noisy conditions. To address this limitation, we propose a DL-based Doppler estimation approach that replaces the ESPRIT processing with a neural network that learns the mapping between received observations and Doppler velocities. Once trained, the model performs estimation through a single forward pass, avoiding iterative computations and enabling efficient inference.

A. Model Description

The proposed estimator employs a multilayer perceptron (MLP) comprising fully connected layers. The input to the network is the received complex baseband signal collected over multiple packets. Since the network operates on real-valued inputs, the complex signal is decomposed into its real and imaginary components, which are then interleaved to form a real-valued input vector. The input vector is processed through a series of fully connected layers that progressively extract features relevant to Doppler velocity estimation. Each layer consists of a linear transformation followed by Layer Normalization (LN) and a ReLU activation function. The final layer outputs the estimated Doppler velocities. Its number of neurons is set equal to the number of targets, (T), enabling simultaneous estimation of the velocities of multiple targets.

B. Training

The MLP model is trained in a supervised manner, where each sample consists of a received complex signal and the corresponding ground truth Doppler velocities. The dataset contains 49,500 samples, with 45,000 used for training and validation (90:10 split) and 4,500 for testing. It is trained by minimizing the Root Mean Squared Error (RMSE) using the Adam optimizer with a weight decay of 10^{-4}, and mini-batch gradient descent with a batch size of 2048. A learning rate scheduler is used with an initial learning rate of 10^{-3}, a patience of 5 epochs, and a minimum learning rate of 10^{-5}. The model with the lowest validation loss is selected. Training is performed offline, and the model is then used for inference.

C. Ablation Study and Architecture Selection

An ablation study is conducted to select the final MLP architecture by evaluating models with different depths and hidden layer sizes for 20 slow time packets and two targets. Architecture I is a two-layer MLP with 100 neurons per layer, Architecture II increases the hidden size to 400 neurons, and Architecture III uses three layers with 400 neurons per layer. All layers use LN and ReLU. A variant of Architecture III without LN is also evaluated. Fig. 2 shows the RMSE performance for models trained at 20 dB SNR and evaluated over [0, 20] dB, and Table I summarizes the complexity and execution time on an embedded processing system. Increasing the hidden size from Architecture I to II reduces RMSE by 7.4%, with 4× more parameters and floating-point operations (FLOPs) and 1.4× higher latency. Increasing depth to Architecture III provides an additional 23.1% reduction in RMSE, with nearly 10× higher complexity and 1.5× higher latency. Overall, Architecture III achieves 30% lower RMSE than Architecture I, with 40× higher complexity and only 2× higher latency. Removing LN increases RMSE by 24.8% with negligible reduction in complexity. Hence, Architecture III with LN is selected and is used for subsequent discussions.

TABLE I: Ablation Study

Architecture	Parameters	FLOPs	PS time (ms)	Model size (KB)
I	4,502	4,600	6.11	20
II	18,002	18,400	8.72	73
III	179,202	180,000	12.85	703
III (w/o LN)	177,602	176,800	12.74	696

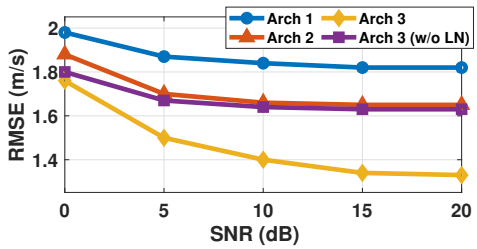

Fig. 2: RMSE comparison of candidate architectures

IV. Hardware Architecture for ESPRIT and MLP on SoC

We present the hardware architectures for both classical ESPRIT and the proposed end-to-end MLP on the Zynq MPSoC via HSCD. Zynq MPSoC comprises a processing system (PS) and FPGA based programmable logic (PL), enabling efficient task partitioning based on computational requirements.

A. ESPRIT

The state-of-the-art hardware architecture of ESPRIT, including all sub-blocks, SS, EVD, and pseudo inverse, is discussed in [5]. In this work, we focus primarily on EVD using an iterative QR decomposition approach. As discussed in section II, in the iterative QR algorithm, the iterations are continued until the termination condition is satisfied. This leads to an unpredictable number of iterations, making it difficult to quantify the real-time complexity of ESPRIT. We thus hard fix the number of iterations, K. Table II shows the ESPRIT RMSE upon fixing K to different values across SNRs. Here, EVD from the PyTorch library is used as a golden reference. Since both accuracy and computational cost increase with K, we select $K = 20$ as the optimal value, ensuring the RMSE is sufficiently close to the golden reference, within 0.7% across all SNRs. Thus, by setting K to 20, we can ensure deterministic latency for ESPRIT in hardware.

In each k^{th} iteration, a series of complex Givens rotations is performed across each column of $\mathbf{A_k}$ to convert it to an upper triangular matrix. These Givens rotations are implemented in a fully pipelined manner. The matrices are partitioned into multiples BRAMs to enable simultaneous MAC operations and memory updates for matrix multiplication operations. All complex multiplications are mapped to DSP blocks for high throughput. Finally, the signal subspace is extracted from the accumulated unitary matrix $\mathbf{Q}$ after K iterations.

TABLE II: PyTorch vs Iterative QR EVD (50 packets)

EVD	K	RMSE (m/s)		
		-20 dB	0 dB	20 dB
PyTorch	-	8.28	2.86	1.09
Iterative QR	1	13.56	5.97	2.48
	10	8.44	2.99	1.21
	20	8.30	2.88	1.09
	30	8.28	2.87	1.09

B. End-to-End MLP and Multi-Model Architecture

The selected MLP model, described in Section III-C, is implemented on a Zynq MPSoC using a hardware software co design approach, where inference is mapped to the PL and the remaining operations are executed on the PS. The design follows a dataflow architecture with all layers instantiated in hardware and connected in a pipelined manner, enabling streaming between layers and improving throughput.

The MLP is realized using a neuron level architecture, where each neuron includes weight memory, bias registers, and a Multiply Accumulate (MAC) unit to compute the weighted sum. Multiple neurons are processed in parallel, controlled by a *parallelism_factor* I that maps L logical neurons to L/I physical neurons. Three configurations are considered, corresponding to different latency and complexity trade offs. Each layer is followed by LN and ReLU, both implemented in hardware. To ensure robust performance across varying conditions, a multi-model approach is adopted. Here, multiple pre-trained MLP model weights are stored in on-chip memory, and a control flag selects the appropriate weights at runtime. Thus, a shared computational architecture is used for all models, avoiding additional hardware overhead, with only increased memory required for storing multiple weight sets. Fig. 3 presents the HSCD of the multi-model approach on Zynq MPSoC.

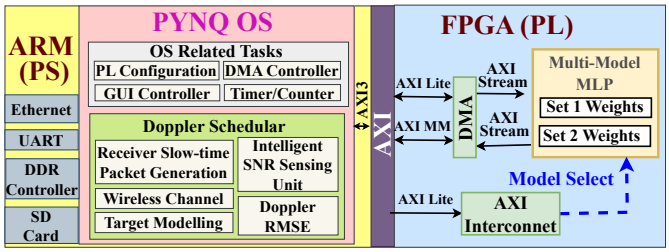

Fig. 3: HSCD of multi-model MLP on Zynq MPSoC

V. Performance Analysis

In this section, the performance of the proposed MLP-based Doppler estimator is compared with the classical ESPRIT algorithm under varying operating conditions, including SNR levels, number of targets, and packet counts. RMSE between the estimated and ground truth Doppler velocities is used as the performance metric. To ensure a fair comparison, the same set of 4,500 ground truth samples is used across all test conditions. Each sample is modeled with $T = 2$ targets, where the range, azimuth, and Doppler velocity are drawn from uniform distributions. The target radar cross-section (RCS) follows a Swirling-1 exponential distribution with a mean of 10 m^2. The ground truth Doppler velocities span $(-30, 30)$ m/s. The center frequency is 60 GHz and the PRI is 2 μs. AWGN is added to the target-reflected returns according to the specified SNR. Each sample first undergoes range-azimuth processing, and the resulting slow-time vector, shown in (1), is used as input to ESPRIT and the proposed MLP. For each operating point, only the noise and observation conditions are varied, while the underlying ground truth remains unchanged.

TABLE III: OOR% comparison between MLP and ESPRIT

Algorithm	N	OOR (%)		
		-20 dB	0 dB	20 dB
MLP	20	0.28	0.25	0.2
ESPRIT	20	81.04	39.03	14.58
	50	37.93	13.33	4.6

A. Comparison of classical ESPRIT and MLP

We begin with highlighting out-of-range (OOR) predictions issue in Doppler velocity estimation. ESPRIT occasionally produces large estimation errors, resulting in predicted velocities that fall far outside the valid Doppler range of $(-30, 30) m/s$. Such outliers can dominate error statistics and lead to misleading performance assessments. To ensure a fair comparison, ESPRIT predictions are constrained within the predefined range by clipping any OOR values to the nearest boundary. As shown in Table III, OOR predictions in MLP is approximately 0.28% and 0.2%, compared to 81.04% and 14.58% for ESPRIT with 20 packets at -20 dB and 20 dB SNR, respectively. This demonstrates that the MLP rarely produces OOR estimates and implicitly learns the valid velocity range making it more reliable and robust.

Next, we evaluate the number of slow-time packets required individually by ESPRIT and MLP to obtain the same RMSE. Fig. 4 shows the RMSE of the MLP with 20 slow-time packets, shown in a red dotted line. We run a number of packet sweep (ranging from 0 to 200) on ESPRIT. Fig. 4(a) shows that at -20 dB SNR, ESPRIT requires 170 packets to achieve the same performance as MLP with 20 packets. Similarly, for an SNR of 20 dB, 40 packets are required by ESPRIT to match the performance of 20 dB MLP as shown in Fig. 4(b). Thus, MLP outperforms ESPRIT, achieving lower RMSE even with fewer packets. Next, Fig. 5 shows the RMSE comparison across SNRs validating superiority of MLP.

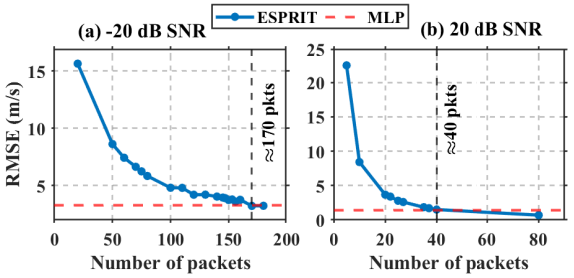

Fig. 4: RMSE comparison between 20-packet MLP and ESPRIT across different packet counts.

B. Effect of training SNR

We evaluate the impact of training SNR on the performance of the selected 20-packet MLP model. Fig. 6 shows the RMSE of models trained at -20 dB, 0 dB, and 20 dB, evaluated over the SNR range of [-20, 20] dB. Each model performs best near its training SNR, with performance degrading as the operating SNR moves away from it. The model trained at -20 dB provides robust performance under low-SNR conditions and remains stable up to approximately -10 dB, beyond which its improvement saturates. The 0 dB model performs well across moderate to high SNRs and achieves a lower error floor. In contrast, the 20 dB model offers no additional gain over the 0 dB model at high SNR while performing worse at low SNR. Based on these observations, only two models are sufficient to cover the entire SNR range effectively. The -20 dB model captures low-SNR conditions, while the 0 dB model generalizes well across moderate and high SNRs, eliminating the need for separate models at each SNR point. The resulting multi-model adaptive approach, shown in Fig. 6, achieves superior performance across all SNRs with minimal complexity overhead, providing up to 56% lower RMSE than a single-model MLP at -20 dB SNR.

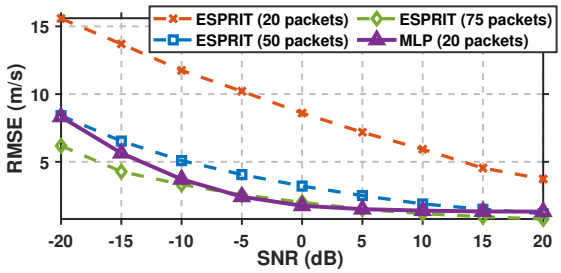

Fig. 5: RMSE comparison between ESPRIT and 20 dB SNR-trained MLP for the localization of two targets.

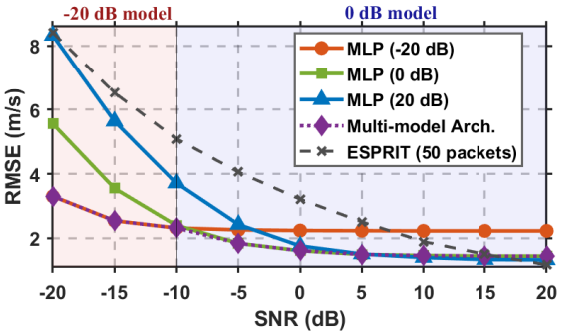

Fig. 6: MLP performance with different training SNRs, and the multi-model architecture.

We next evaluate the proposed multi-model architecture for varying numbers of targets. Fig. 7 compares the performance of the model trained and tested with $T = 3$ against classical ESPRIT. While ESPRIT with 50 packets suffers significant performance degradation as the number of targets increases, the proposed architecture with 20 packets clearly outperforms ESPRIT with 50 packets and even surpasses ESPRIT with 75 packets at low and moderate SNRs. This demonstrates that the proposed end-to-end MLP approach is more robust to target variations and provides better resolution than classical ESPRIT.

VI. Hardware Complexity Analysis

This section compares the hardware complexity of ESPRIT, end-to-end MLP, and multi-model MLP on a Zynq MPSoC platform. The designs are implemented using double-precision

TABLE IV: Hardware complexity comparison of Doppler estimation architectures on Zynq MPSoC.

Architecture	N	Exec. Time (ms)		AF w.r.t	Resource Utilization			Power
		PS	PL	ESPRIT	BRAM	DSP	LUT	(W)
ESPRIT	50	1257.28	7.15	175.8	59	978	118479	5.77
High Lat. MLP	20	12.85	0.56	2246.2	178 (+201.7%)	97 (-90.1%)	129700 (+9.5%)	4.79
Med. Lat. MLP		12.85	0.43	2910.4	192 (+225.4%)	167 (-82.9%)	131406 (+10.9%)	4.91
Low Lat. MLP		12.85	0.35	3592.2	212 (+259.3%)	695 (-28.9%)	186685 (+57.6%)	6.03
Multi-model		12.85	0.36	3492.4	415 (+603.4%)	699 (-28.5%)	175008 (+47.7%)	5.85

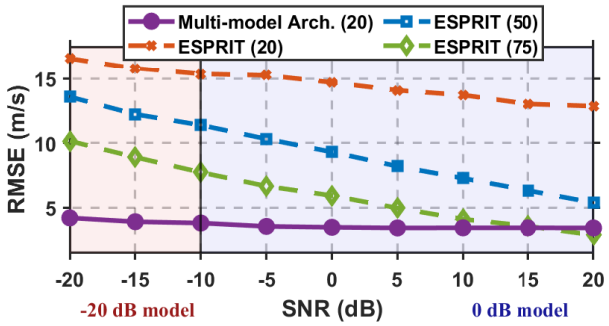

Fig. 7: RMSE comparison between ESPRIT and multi-model MLP for localization of three targets.

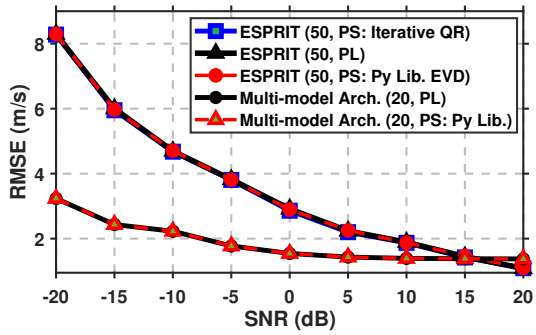

Fig. 8: PS and PL RMSE comparison for ESPRIT and MLP.

(DPFL) and single-precision (SPFL) floating-point on the PS and PL, respectively. Functional accuracy is validated by comparing PS and PL implementations. As shown in Fig. 8, the curves closely overlap for both ESPRIT and MLP, confirming that the SPFL hardware maintains accuracy relative to DPFL implementations. In addition, the 20-packet-multi-model MLP achieves a 59.8% RMSE improvement over 50-packet ESPRIT at -20 dB SNR.

Table IV summarizes the hardware complexity and acceleration factor (AF), computed with respect to the PS execution time of 50-packet ESPRIT. All MLP configurations achieve significant latency reduction, providing 12.8–20.4× speedup over ESPRIT. The high, medium, and low latency designs correspond to different parallelism factors, $I = 20, 20, 1$, $10, 10, 1$, and $2, 2, 1$, respectively. Increasing the degree of parallelism reduces execution time, with up to 1.6× improvement, but leads to a proportional increase in DSP utilization, reflecting the trade off between latency and hardware resources.

The multi-model architecture, implemented with $I = 2, 2, 1$, introduces only memory overhead due to storage of multiple weight sets, while maintaining similar DSP usage and negligible latency increase (0.35 ms to 0.36 ms). Overall, while ESPRIT uses fewer BRAM and LUT resources, it exhibits significantly higher latency and DSP usage, along with inferior estimation performance. These results demonstrate the superiority of MLP over subspace methods.

VII. Conclusion

This work demonstrates that MLP-based Doppler velocity estimation is an effective alternative to classical subspace methods for mmW ISAC systems. The proposed end-to-end MLP improves estimation accuracy by 59.8% and reduces Doppler estimation latency by 20.4× over classical ESPRIT, while reducing reliance on large packet sizes and enabling real-time operation. A simple multi-model approach further ensures robust performance across varying SNR conditions using only two models, eliminating the need for specialized designs. Overall, the proposed method achieves a favorable trade-off between accuracy, complexity, and adaptability, making it well suited for resource-constrained ISAC systems.

References

[1] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," IEEE Transactions on Communications, vol. 68, no. 6, pp. 3834–3862, 2020.

[2] K. V. Mishra, M. Bhavani Shankar, V. Koivunen, B. Ottersten, and S. A. Vorobyov, "Toward millimeter-wave joint radar communications: A signal processing perspective," IEEE Signal Processing Magazine, vol. 36, no. 5, pp. 100–114, Sep. 2019.

[3] A. Sneh, S. Jain, V. S. Sindhu, S. S. Ram, and S. Darak, "Ieee 802.11ad based joint radar communication transceiver: Design, prototype and performance analysis," IEEE Transactions on Vehicular Technology, pp. 1–16, 2023.

[4] A. Tewari, S. S. Jha, A. Sneh, S. J. Darak, and S. S. Ram, "Reconfigurable radar signal processing accelerator for integrated sensing and communication system," IEEE Transactions on Aerospace and Electronic Systems, pp. 1–19, 2024.

[5] A. Tewari, S. S. Bhardwaj, S. J. Darak, and S. S. Ram, "Reconfigurable low-complexity architecture for high resolution doppler velocity estimation in integrated sensing and communication system," in IEEE Radar Conference, 2026.

[6] X.-W. Zhang, D. Yan, L. Zuo, M. Li, and J.-X. Guo, "High-performance of eigenvalue decomposition on fpga for the doa estimation," IEEE Transactions on Vehicular Technology, vol. 72, no. 5, pp. 5782–5797, 2023.

[7] D. H. Shmuel, J. P. Merkofer, G. Revach, R. J. G. van Sloun, and N. Shlezinger, "Subspacenet: Deep learning-aided subspace methods for doa estimation," IEEE Transactions on Vehicular Technology, vol. 74, no. 3, pp. 4962–4976, 2025.

[8] A. Gast, L. Le Magoarou, and N. Shlezinger, "Dcd-music: Deep-learning-aided cascaded differentiable music algorithm for near-field localization of multiple sources," in ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2025, pp. 1–5.