WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal Homepage: www.ijrpr.com ISSN 2582-7421

Iterative Structured Support Detection with Ridge-Regularized Fusion for Low-Overhead Beamspace Channel Estimation in Lens-Array mmWave Massive MIMO

Puja Kumari[1], Dr. Pramod Sharma[2]

[1] M.Tech Scholar (Digital Communication), Regional College for Education Research & Technology, Jaipur, Rajasthan, India
[2] Professor, Regional College for Education Research & Technology, Corresponding author: Pramod.Sharma@deepshikha.org

ABSTRACT—

Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) systems equipped with a lens antenna array convert the spatial channel into a sparse beamspace channel, so only a small number of radio-frequency (RF) chains is needed. The price is that the base station (BS) observes a high-dimensional channel through very few RF chains, which makes channel estimation an underdetermined sparse-recovery problem. Support-detection (SD) estimators that exploit the structure of the beamspace channel outperform generic compressed-sensing (CS) solvers. However, they use a fixed support window, detect each path only once and fuse the components by plain least squares (LS), so they show an error floor at high signal-to-noise ratio (SNR) and are sensitive to error propagation at low SNR. This paper proposes an Iterative Structured Support Detection with Ridge fusion (ISD-R) estimator that adds three elements: (i) an SNR-adaptive support window, (ii) back-and-forth successive-interference-cancellation (SIC) re-detection of every path position, and (iii) ridge-regularized joint fusion. With N = 256 antennas, K = 16 users and Q = 96 pilot instants, ISD-R lowers the normalized mean-square error (NMSE) by 2.2 dB compared with SD at 0 dB SNR and by 1.2 dB at 30 dB. It outperforms orthogonal matching pursuit (OMP) by 2.8–6.2 dB and runs about 2.5× faster than OMP. With N = 512, the high-SNR gain over SD grows to 2.5 dB. We also report the operating points where ISD-R gives no gain.

***Index Terms*—Massive MIMO, millimeter-wave, lens antenna array, beamspace channel estimation, support detection, compressed sensing, ridge regression, pilot overhead.**

I. INTRODUCTION

Massive MIMO, which deploys tens to hundreds of antennas at the BS, is a key physical-layer technology of 5G and beyond-5G networks [1], [2]. Combined with the wide bandwidths of the mmWave bands (28–100 GHz), it can provide multi-gigabit user rates. However, mmWave propagation suffers from high path loss, so very large arrays and highly directional beamforming are required [3]. A fully digital array needs one RF chain per antenna, which is too costly and power-hungry for arrays with hundreds of elements [3], [4].

Beamspace MIMO using a lens antenna array solves this hardware problem [5]. The lens acts as a spatial discrete Fourier transform (DFT) and focuses the energy of each propagation path onto a few beams. Because mmWave channels have only a few dominant paths, the beamspace channel is sparse, and beam selection allows the BS to use only as many RF chains as there are users [6]. Beam selection, however, needs accurate knowledge of the beamspace channel, which the BS sees through only $N_{\mathrm{RF}} \ll N$ RF chains. Estimating it from short pilot sequences is therefore underdetermined.

Exhaustive beam sweeping (the SMD benchmark) needs $Q \geq N$ pilot instants. CS solvers such as OMP [7] and CoSaMP [8] reduce the overhead, but they treat all nonzero entries as independent unknowns and often find the wrong support at low uplink SNR. Gao et al. [9] observed that each path occupies a contiguous cluster of strong beams around its peak. Their SD estimator detects one peak at a time, keeps a fixed window of V beams around it, and removes the estimated path before searching for the next one. Deep-learning estimators [10], [11] have also been proposed, but they need training data and must be retrained when the configuration changes.

The SD estimator has three weaknesses:

1) The window size V is fixed. At high SNR, the leakage outside the window causes an NMSE floor, and at low SNR a large window adds noise.
2) Each path is detected once, in greedy order, so an early error carries over to later steps.
3) The final LS fusion amplifies noise when the columns of the measurement matrix are correlated.

Contributions. The contributions of this paper are:

- An ISD-R estimator that combines an SNR-adaptive structured window, back-and-forth SIC re-detection and ridge-regularized fusion.
- A reproducible evaluation with the system parameters of [9]. ISD-R is compared with SMD, OMP and SD in two array configurations, using NMSE versus SNR, NMSE versus pilot overhead and measured run time.
- An honest operating-region analysis that identifies where the refinements help and where they do not.

This work extends the M.Tech thesis research of the first author on channel estimation and pilot decontamination for 5G massive MIMO.

II. RELATED WORK

Beamspace architectures. Brady et al. [5] built and measured a lens-based beamspace MIMO prototype, which confirmed that a discrete lens array behaves as a spatial DFT. Amadori and Masouros [6] proposed beam-selection rules that reduce the number of RF chains to about the number of users with little loss in sum rate. Both studies assume perfect channel state information (CSI).

TABLE I — Qualitative Comparison with Related Estimators

Method	**Struct. prior**	**Adapt. window**	**Re-det.**	**Regul. fusion**	**Training data**
OMP [7]	–	–	–	–	No
CoSaMP [8]	–	–	partial	–	No
Adaptive CS [4]	codebook	–	–	–	No
SD [9]	yes	–	–	–	No
LDAMP [10]	learned	–	iter.	learned	Yes
DL-sparse [11]	learned	–	–	learned	Yes
ISD-R	**yes**	**yes**	**yes**	**ridge**	**No**

CS estimators. Alkhateeb et al. [4] formulated channel estimation for hybrid mmWave architectures as an adaptive CS problem with multiresolution codebooks. OMP [7] and CoSaMP [8] are the standard greedy baselines. Neither exploits the clustered structure of the beamspace.

Structure-aware and learned estimators. Gao et al. [9] introduced the adaptive selecting network and the SD estimator. The thesis underlying this paper reproduced SD and confirmed a pilot saving of about 50% compared with OMP. He et al. [10] unfolded approximate message passing into a learned denoising network, and Ma et al. [11] learned the measurement matrix and the estimator jointly. Both need large training sets. Table I positions ISD-R among these methods.

III. SYSTEM MODEL

A. mmWave Massive MIMO with Lens Antenna Array

We consider a time-division duplex (TDD) system in which a BS with an N-element uniform linear array (ULA) and $N_{\mathrm{RF}} = K$ RF chains serves K single-antenna users. Following the Saleh–Valenzuela model, the channel of user k is

$$\mathbf{h}_k = \sqrt{(N/(L_k+1))}\ \Sigma_{i=0}^{Lk}\}\ \beta_k^{(i)}\ \mathbf{a}(\psi_k^{(i)}), \qquad (1)$$

where $i = 0$ denotes the line-of-sight (LoS) path, L_k is the number of non-line-of-sight (NLoS) paths, $\beta_k^{(i)}$ is the complex gain and $\psi_k^{(i)} = (d/\lambda)\sin\theta_k^{(i)}$ is the spatial direction. The ULA response vector is

$$\mathbf{a}(\psi) = (1/\sqrt{N})\ [\ e^{-j2\pi\psi m}\ ]_{m\in\mathcal{J}(N)}, \qquad (2)$$

where $\mathcal{J}(N) = \{p - (N-1)/2 : p = 0, \ldots, N-1\}$. The lens implements the spatial DFT matrix $\mathbf{U} = [\mathbf{a}(\bar{\psi_1}), \ldots, \mathbf{a}(\bar{\psi_N})]^{\mathrm{H}}$ with $\bar{\psi_n} = (1/N)(n - (N+1)/2)$, so the beamspace channel is

$$\tilde{\mathbf{H}} = \mathbf{UH} = [\tilde{\mathbf{h}}_1, \ldots, \tilde{\mathbf{h}}_K]. \qquad (3)$$

Fig. 1 shows the architecture.

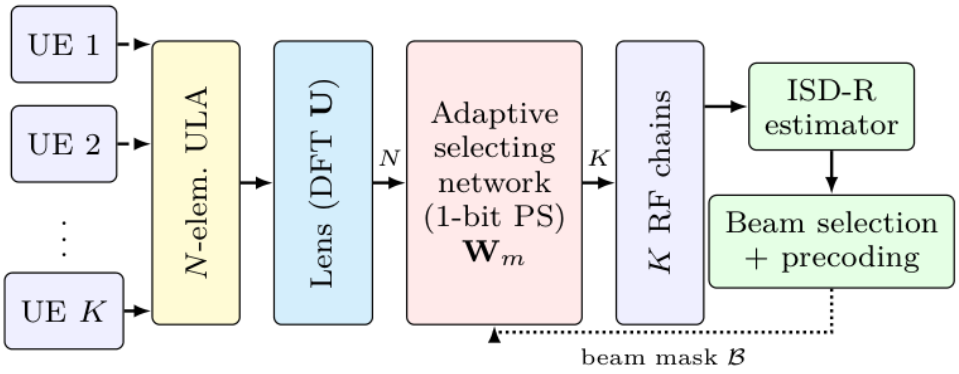


Fig. 1. Lens-array mmWave massive MIMO uplink training architecture with an adaptive selecting network and the proposed ISD-R estimator.

B. Pilot Transmission and Measurement Model

Training uses $Q = MK$ instants divided into M blocks. In block m the users transmit orthogonal pilots $\mathbf{\Psi}_m \in \mathbb{C}^{K\times K}$ with $\mathbf{\Psi}_m\mathbf{\Psi}_m^{\mathrm{H}} = \mathbf{I}_K$, and the BS combines the received signals with $\mathbf{W}_m \in \mathbb{C}^{K\times N}$, which is realized by the adaptive selecting network [9]. After decorrelating the pilots, the BS obtains

$$\mathbf{Z}_m = \mathbf{W}_m\tilde{\mathbf{H}} + \mathbf{W}_m\mathbf{N}_m\mathbf{\Psi}_m^{\mathrm{H}}. \qquad (4)$$

Stacking the kth columns over all M blocks gives

$$\bar{\mathbf{z}}_k = \bar{\mathbf{W}}\,\tilde{\mathbf{h}}_k + \bar{\mathbf{n}}_k,\ \ \bar{\mathbf{W}} \in \mathbb{C}^{Q\times N},\ \ Q \ll N. \qquad (5)$$

Each phase shifter takes values in $\{\pm 1/\sqrt{N}\}$, so we model the entries of $\bar{\mathbf{W}}$ as independent Bernoulli $\pm 1/\sqrt{N}$ variables [9]. The uplink SNR is $1/\sigma^2_{\mathrm{UL}}$.

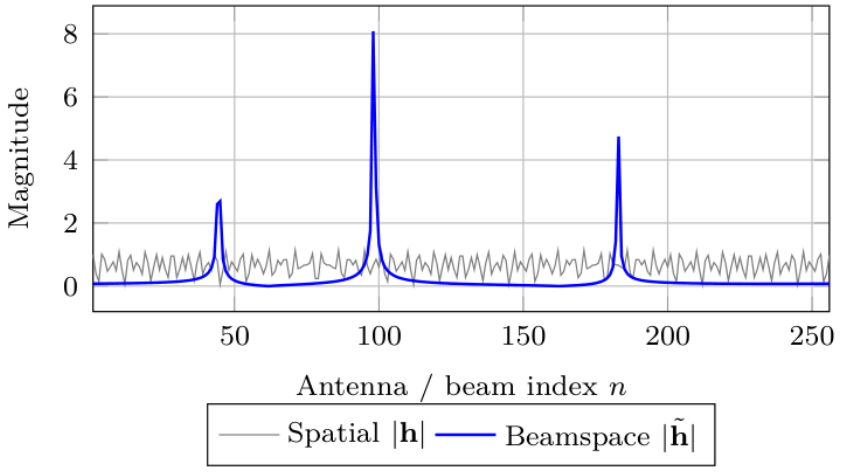


Fig. 2. One user's channel (N = 256, 1 LoS + 2 NLoS paths). The spatial channel is spread over all antennas, whereas the beamspace channel concentrates each path in a narrow cluster of beams.

C. Structural Sparsity

Each path contributes a Dirichlet-kernel profile to $\tilde{\mathbf{h}}_k$. The profile peaks at the beam nearest to $\psi_k^{(i)}$ and decays as $1/|n - n^*|$, so a window of $V = 8$ contiguous beams captures more than 95% of the path energy [9]. Each path is therefore described by a single position parameter. Fig. 2 illustrates this property.

IV. PROPOSED ISD-R ESTIMATOR

A. Stage 1: SNR-Adaptive Structured Window

The window size controls a bias–variance tradeoff. A small window discards the leakage tails (bias, which causes the high-SNR floor), and a large window adds unknowns that must be estimated from noisy data (variance). ISD-R chooses V from the SNR estimate $\hat{\rho}$:

$$V(\rho) = \begin{cases} 8, & \hat{\rho} < 17.5 \text{ dB} \\ 12, & 17.5 \le \hat{\rho} < 25 \text{ dB} \\ 16, & \hat{\rho} \ge 25 \text{ dB} \end{cases} \quad (6)$$

subject to $V(L+1) \le Q$.

B. Stage 2: Greedy Initialization

For $l = 1, \ldots, L+1$, the estimator takes the correlation peak $p_l = \arg\max_n |[\bar{\mathbf{W}}^{\mathrm{H}}\mathbf{r}]_n|$, forms the window $S_l = \{p_l - V/2, \ldots, p_l + V/2 - 1\}$ (indices modulo N), estimates the component on S_l by ridge LS and subtracts it from the residual $\mathbf{r}$. With $V = 8$ and $\mu = 0$, this stage is identical to SD [9].

Algorithm 1 ISD-R beamspace channel estimation (user k)

Require: $\bar{\mathbf{z}}_k$, $\bar{\mathbf{W}}$, L, σ^2, T_{p}

1: Select V by (6); $\mu \leftarrow \sigma^2/2$; $\mathbf{r} \leftarrow \bar{\mathbf{z}}_k$

2: **for** $l = 1$ **to** $L+1$ **do**

3: $p_l \leftarrow \arg\max_n |[\bar{\mathbf{W}}^{\mathrm{H}}\mathbf{r}]_n|$; form S_l

4: $\hat{\mathbf{x}} \leftarrow$ ridge LS of $\mathbf{r}$ on S_l; $\mathbf{r} \leftarrow \mathbf{r} - \bar{\mathbf{W}}\hat{\mathbf{x}}$

5: **end for**

6: **for** $t = 1$ **to** T_{p} **do**

7: **for** $l = 1$ **to** $L+1$ **do**

8: $\mathbf{r}_l \leftarrow \bar{\mathbf{z}}_k - \bar{\mathbf{W}}\,\mathrm{ridge}(\bar{\mathbf{z}}_k; \cup_{j \ne l} S_j)$

9: $p_l \leftarrow \arg\max_n |[\bar{\mathbf{W}}^{\mathrm{H}}\mathbf{r}_l]_n|$; update S_l

10: **end for**

11: **end for**

12: Compute $\tilde{\hat{\mathbf{h}}}_k$ on $S = \cup_l S_l$ by (8)

13: **return** $\tilde{\hat{\mathbf{h}}}_k$

C. Stage 3: Back-and-Forth SIC Re-detection

The first peak p_1 is detected while the other paths are still present, so their leakage can shift it. ISD-R therefore runs T_{p} refinement passes. In each pass, for every l, it jointly re-estimates all other components and removes them from the original measurement:

$$\mathbf{r}_l = \bar{\mathbf{z}}_k - \bar{\mathbf{W}}\,\hat{\mathbf{x}}_{\cup_{\{j \ne l} S_j\}}. \quad (7)$$

It then re-detects $p_l = \arg\max_n |[\bar{\mathbf{W}}^{\mathrm{H}}\mathbf{r}_l]_n|$. This is a coordinate-descent search over the $L+1$ position parameters, similar in spirit to space-alternating expectation-maximization, with path positions as the unknowns.

D. Stage 4: Ridge-Regularized Joint Fusion

With $S = \cup_l S_l$, the final estimate is

$$\tilde{\hat{\mathbf{h}}}_{k,S} = (\bar{\mathbf{W}}_S^{\mathrm{H}}\bar{\mathbf{W}}_S + \mu\mathbf{I})^{-1}\,\bar{\mathbf{W}}_S^{\mathrm{H}}\,\bar{\mathbf{z}}_k, \;\; \mu = \tfrac{1}{2}\,\sigma^2, \qquad (8)$$

and all entries outside S are set to zero. Ridge fusion is the linear MMSE estimator under a unit-variance prior on the coefficients. It bounds the noise gain when $\bar{\mathbf{W}}_S$ is poorly conditioned, for example when windows overlap or when $V(L+1)$ approaches Q. Algorithm 1 and Fig. 3 summarize the procedure.

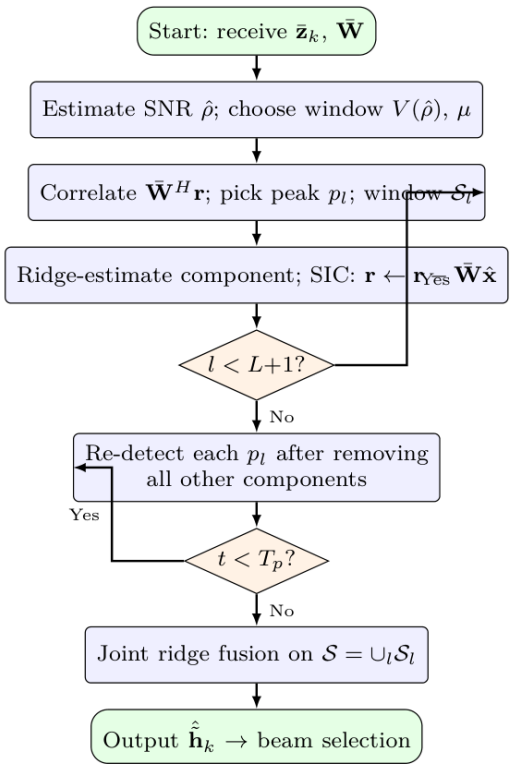


Fig. 3. Flowchart of the proposed ISD-R estimator.

E. Complexity

OMP with sparsity $s = V(L+1)$ costs about $O(sQN + s^3Q)$, and SD costs $O((L+1)(QN + QV^2))$. ISD-R costs $O((1+T_p)(L+1)QN + T_p(L+1)Q(LV)^2)$. Because $(1+T_p)(L+1) = 9 < s = 24$, ISD-R is cheaper than OMP, as the run times in Table IV confirm.

F. Hardware Specification

Table II compares the front-end component counts. The adaptive selecting network uses only 1-bit phase shifters (sign switches), and the same network also acts as the combiner during training.

TABLE II — Front-End Hardware (N = 256, K = 16; PS: phase shifter; ASN: adaptive selecting network)

Architecture	**RF chains**	**PS / switches**	**PS res.**
Fully digital	256	0	–
Hybrid, fully conn.	16	4096 PS	4–6 b
Lens + switches	16	4096 sw.	–
Lens + ASN (ours)	**16**	**4096 PS**	**1 b**

V. SIMULATION SETUP

The parameters follow [9] and the thesis work and are listed in Table III. The NMSE is defined as $\mathbb{E}\|\hat{\tilde{\mathbf{h}}} - \tilde{\mathbf{h}}\|^2 / \mathbb{E}\|\tilde{\mathbf{h}}\|^2$. All estimators share the same combiner and noise realizations. The simulator was written in Python/NumPy, and run times were measured on a single CPU core.

TABLE III — Simulation Parameters

Parameter	**Value**
N / N_{RF} / K (Case 1)	256 / 16 / 16
N / N_{RF} / K (Case 2)	512 / 32 / 32

Parameter	**Value**
Channel model	S–V, 1 LoS + L_k=2 NLoS
Path gains	CN(0, 1), CN(0, $10^{-0.5}$)
Spatial directions	i.i.d. U[−0.5, 0.5]
Pilot instants Q	96 (Case 1), 192 (Case 2)
Combiner	random 1-bit ($\pm 1/\sqrt{N}$)
SD window / OMP sparsity	V = 8 / 24
ISD-R	$V \in \{8, 12, 16\}$, $T_p = 2$, $\mu = \sigma^2/2$
Carrier / bandwidth	28 GHz / 1 GHz
Monte-Carlo trials	150 (Case 1) / 60 (Case 2)

VI. RESULTS AND DISCUSSION

A. NMSE versus SNR

Fig. 4 shows the NMSE as a function of SNR for Case 1. OMP performs worst, and below 5 dB it is even worse than SMD, because it often selects the wrong support. SD gains 3.0–4.3 dB over OMP between 0 and 10 dB, but it saturates near −15 dB because of leakage outside its fixed windows. ISD-R improves on SD in two regions:

- Low SNR (0 dB): −2.70 dB versus −0.49 dB, a 2.2 dB gain, because re-detection and ridge shrinkage reduce error propagation and noise amplification.
- High SNR (30 dB): −16.27 dB versus −15.05 dB, a 1.2 dB gain, because the wider window captures the leakage tails.

At 10–15 dB the two methods agree to within 0.05 dB. SMD keeps improving with SNR, but it needs Q = 256 pilot instants, 2.7 times as many as the other methods.

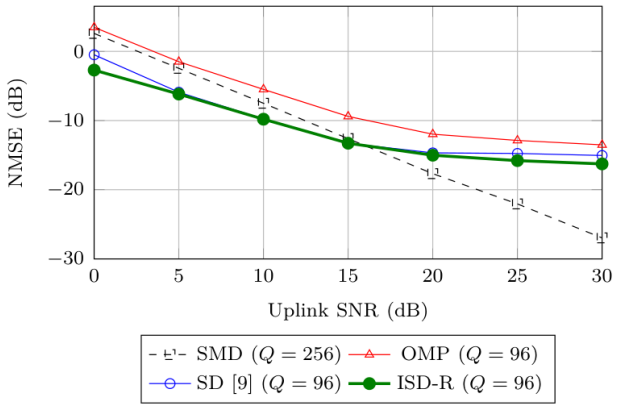


Fig. 4. NMSE versus uplink SNR, Case 1 (N = 256, K = 16).

B. NMSE versus Pilot Overhead

Fig. 5 shows the NMSE as a function of Q at 10 dB SNR. OMP needs about Q = 192 to reach −9 dB, whereas both structured estimators reach this level at $Q \approx 96$, which is a 50% pilot saving and agrees with the thesis results. With short training, ISD-R adds a further gain over SD of 1.1 dB at Q = 48 and 0.6 dB at Q = 64, because the ridge fusion is most useful when the matrix is poorly conditioned. For $Q \geq 96$, SD is marginally better (by 0.03–0.28 dB), because the ridge bias then outweighs the variance reduction. This suggests letting $\mu \to 0$ when $Q \gg V(L+1)$. The SMD curve is shown only as a reference, since SMD always uses a full sweep.

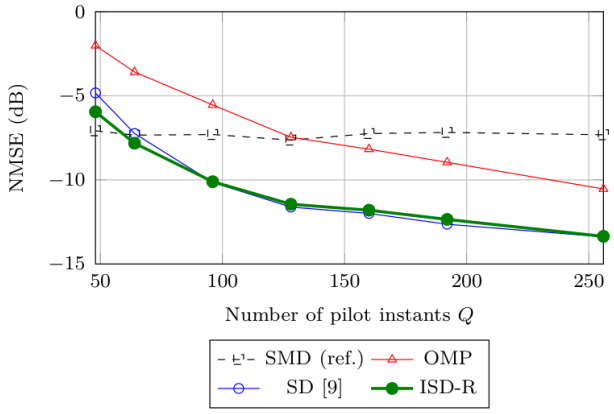


Fig. 5. NMSE versus pilot overhead Q at 10 dB SNR, Case 1.

C. Scalability: Case 2 (N = 512, K = 32)

Fig. 6 shows the results for the larger array. At high SNR, the advantage of ISD-R grows: −17.49 dB versus −14.95 dB for SD at 30 dB (2.5 dB) and −17.72 versus −15.45 dB at 25 dB (2.3 dB). The error floors of OMP and SD even rise slightly above 25 dB. At 0–10 dB, however, SD is better by 0.3–2.0 dB, because with $N = 512$ the threshold rule (6) is not tuned and re-detection can move a weak NLoS peak onto a noise peak. The thresholds should therefore scale with N/Q. A hybrid rule that uses SD below about 15 dB and ISD-R above it would follow the lower envelope of the two curves.

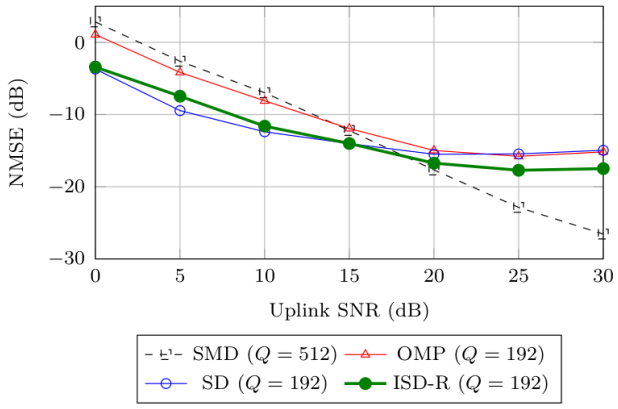


Fig. 6. NMSE versus uplink SNR, Case 2 (N = 512, K = 32).

D. Gain Analysis

Fig. 7 shows that the gain of ISD-R over SD follows a U-shape: it is large at both ends of the SNR range and close to zero in the middle. This matches the bias–variance reasoning in Section V. The gain over OMP is always positive (2.8–6.2 dB) and is largest at low SNR.

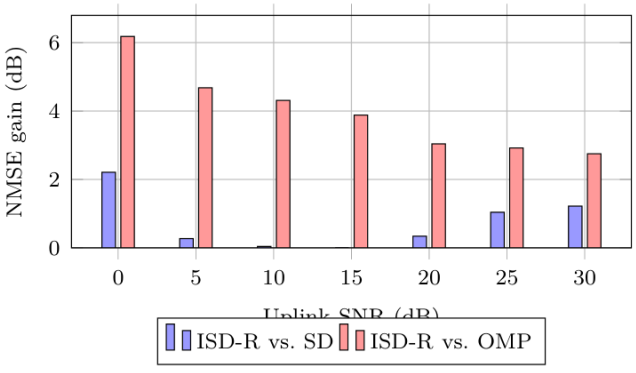


Fig. 7. NMSE gain of ISD-R over SD and OMP, Case 1 (Q = 96).

E. Comparative Summary

Table IV summarizes the results. ISD-R takes about 5.8 times as long as SD to run, but it is 2.5 times faster than OMP and uses 62.5% fewer pilots than SMD.

TABLE IV — Comparison of Estimators (Case 1, N = 256)

Method	**Q**	**NMSE (dB) 0 dB**	**NMSE (dB) 30 dB**	**Time/user (ms)**
SMD (beam sweep)	256	2.60	−26.97	–
OMP [7]	96	3.48	−13.51	5.00
SD [9]	96	−0.49	−15.05	0.35
ISD-R	**96**	**−2.70**	**−16.27**	**2.01**

F. Practical Implementation Notes

SNR estimation. The SNR estimate $\hat{\rho}$ is obtained from the residual after initialization, $\sigma^2 \approx \|\mathbf{r}\|^2/Q$, so no extra pilots are needed.

Parallelism. The estimator runs independently for each user, so the K instances can be executed in parallel.

Latency. The coherence time at 28 GHz and a user speed of 30 m/s is about 0.36 ms. The measured Python run time must therefore be reduced by one to two orders of magnitude through fixed-point hardware, which is realistic because the dominant operations are $Q \times N$ correlations.

Robustness. If the number of paths L is overestimated, ridge fusion shrinks the spurious components.

VII. CONCLUSION

This paper presented ISD-R, an iterative structured support-detection estimator with ridge fusion for lens-array mmWave massive MIMO. Compared with SD, it achieves up to 2.2 dB lower NMSE at low SNR and up to 2.5 dB lower NMSE at high SNR. It outperforms OMP at every evaluated SNR while running 2.5 times faster, and it uses 62.5% fewer pilots than beam sweeping. Future work includes N/Q-aware window adaptation, wideband channels with beam squint, and integration with the pilot-decontamination framework of our companion paper.

REFERENCES

[1] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” IEEE Trans. Wireless Commun., vol. 9, no. 11, pp. 3590–3600, 2010, doi:10.1109/TWC.2010.092810.091092.

[2] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO has unlimited capacity,” IEEE Trans. Wireless Commun., vol. 17, no. 1, pp. 574–590, 2018.

[3] R. W. Heath Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, “An overview of signal processing techniques for millimeter wave MIMO systems,” IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 436–453, 2016.

[4] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath Jr., “Channel estimation and hybrid precoding for millimeter wave cellular systems,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 831–846, 2014.

[5] J. Brady, N. Behdad, and A. M. Sayeed, “Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements,” IEEE Trans. Antennas Propag., vol. 61, no. 7, pp. 3814–3827, 2013, doi:10.1109/TAP.2013.2254442.

[6] P. V. Amadori and C. Masouros, “Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection,” IEEE Trans. Commun., vol. 63, no. 6, pp. 2212–2223, 2015, doi:10.1109/TCOMM.2015.2431266.

[7] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” IEEE Trans. Inf. Theory, vol. 53, no. 12, pp. 4655–4666, 2007.

[8] D. Needell and J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” Appl. Comput. Harmon. Anal., vol. 26, no. 3, pp. 301–321, 2009.

[9] X. Gao, L. Dai, S. Han, C.-L. I, and X. Wang, “Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array,” IEEE Trans. Wireless Commun., vol. 16, no. 9, pp. 6010–6021, 2017, doi:10.1109/TWC.2017.2718502.

[10] H. He, C.-K. Wen, S. Jin, and G. Y. Li, “Deep learning-based channel estimation for beamspace mmWave massive MIMO systems,” IEEE Wireless Commun. Lett., vol. 7, no. 5, pp. 852–855, 2018.

[11] W. Ma, C. Qi, Z. Zhang, and J. Cheng, “Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO,” IEEE Trans. Commun., vol. 68, no. 5, pp. 2838–2849, 2020.