An Enhanced Diffusion Model for (Overcoming Limitations) in mmWave Massive MIMO Systems Channel Estimation

Jinwook Kim, *Student Member, IEEE*, Seongwoo Lee, *Graduate Student Member, IEEE*, Soo Hyun Kim, *Graduate Student Member, IEEE*, Young Ghyu Sun, *Member, IEEE*, Joonho Seon, *Graduate Student Member, IEEE*,Hyowoon Seo, *Member, IEEE*, Dong In Kim, *Life Fellow, IEEE*, and Jin Young Kim, *Senior Member, IEEE*

***Abstract*—mmWave massive MIMO has been a key technology of modern wireless communication systems, which can provide high data rates and high spectral efficiency thanks to the large number of antenna arrays and high carrier frequencies. Acquisition of channel state information by channel estimation must be performed to achieve these gains. Traditional channel estimation approaches such as least squares (LS) and minimum mean squared error (MMSE) have been suffered from degraded performance and the increased number of pilot symbols due to the high dimensionality from the large number of antennas, leading to reduction in spectral efficiency. To address this problem, compressed sensing (CS) based approaches have been proposed to reduce the pilot overhead by leveraging the inherent sparsity of mmWave channels.**

***Index Terms*—Channel estimation, mmWave, massive MIMO, diffusion model, score function, Tweedie's formula.**

I. INTRODUCTION

Millimeter-wave (mmWave) massive multiple-input-multiple-output (MIMO) has been a key technology of modern wireless communication systems, which can provide high data rates and high spectral efficiency thanks to the large number of antenna arrays and high carrier frequencies. Acquisition of channel state information by channel estimation must be performed to achieve these gains. Traditional channel estimation approaches such as least squares (LS) and minimum mean squared error (MMSE) have been suffered from degraded performance and the increased number of pilot symbols due to the high dimensionality from the large number of antennas, leading to reduction in spectral efficiency [1]. To address this problem, compressed sensing (CS) based approaches have been proposed to reduce the pilot overhead by leveraging the inherent sparsity of mmWave channels [2]–[4].

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-00892-005, Research on advanced physical layer technologies of low-earth orbit (LEO) satellite communication systems for ubiquitous intelligence in space) and supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2025-RS-2023-00258639) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

Jinwook Kim, Seongwoo Lee, Soo Hyun Kim, Young Ghyu Sun, Joonho Seon, and Jin Young Kim are with the Department of Electronic Convergence Engineering, Kwangwoon University, Seoul 01897, South Korea (e-mail: {yoonlight12, swoo1467, kimsoogus, yakrkr, dimlight13, jinyoung}@kw.ac.kr).

Hyowoon Seo, and Dong In Kim are with the Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, South Korea (e-mail: {hyowoonseo,dongin}@skku.edu).

But, they have performance limitations because they assume that the channel is mathematically ideally sparse, whereas real-world channels are not. Furthermore, the performance of the model is very sensitive to the sensing matrix and system parameters [4].

Recently, deep learning-based channel estimation has much gained interest to be enable to approximate complex data [5], [6]. Diffusion model is one of the generative models have been. ARVINTE was the first study to use the diffusion model for channel estimation and demonstrated the performance of the model and its performance in out-of-distribution, but it suffers from slow channel estimation due to the very high sampling rate [7]. In order to solve many sampling steps of the diffusion model, the problem was solved with a denoising method based on progressive MSE. The denoising method requires an LS estimate, and this method requires the same number of pilot symbols and transmit antennas, which can cause pilot overhead problems, especially in downlink communication [8]. The zhou paper improved the sampling rate by assuming a uniform prior and showed good performance at LOS. However, the performance in NLOS environment with low pilot overhead was not investigated. Furthermore, it is characterized by the assumption of variance [9].

In this letter, the novel channel estimation algorithm has been proposed to handle the accuracy, pilot overhead, and complexity simultaneously.

II. SYSTEM MODEL

In this letter, mmWave massive MIMO downlink channel is considered. To simplify, uniform linear array with half wavelength and a quasi-static channel is assumed. Received pilot symbol is expressed as,

$$\mathbf{Y} = \mathbf{H}\mathbf{P} + \mathbf{N} \in \mathbb{C}^{N_r \times N_p} \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is channel matrix, $\mathbf{P} \in \mathbb{C}^{N_t \times N_p}$ is known pilot symbol, and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ is additive white Gaussian noise. Under the assumption of ULA and channel matrix can be expressed as virtual channel representation as

$$\mathbf{H} = \mathbf{A}_R \mathbf{H}_V \mathbf{A}_T^H, \tag{2}$$

where $\mathbf{A}_R$ and $\mathbf{A}_T$ are DFT matrices and $\mathbf{H}_V$ is channel matrix of angular domain. Vectorized form of received symbol is expressed as,

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{n} \in \mathbb{C}^{N_r N_p \times 1}, \tag{3}$$

where $\mathbf{y}$, $\mathbf{h}_V$, $\mathbf{n}$ are vector form of received symbol, channel matrix of angular domain, and noise each. $\mathbf{A} = (\mathbf{P}^\top \otimes \mathbf{I}_{N_r})((\mathbf{A}_T^H)^\top \otimes \mathbf{A}_R) \in \mathbb{C}^{N_r N_p \times N_r N_t}$ is vectorized by Kronecker product.

III. Proposed Method

$$\mathbf{h}_t = \sqrt{\bar{\alpha}_t}\mathbf{h}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t, \tag{4}$$

$$\mathcal{L}(\theta) = \mathbb{E}[\|\epsilon_t - \epsilon_\theta(\mathbf{h}_t, t)\|_2^2], \tag{5}$$

$$\mathbf{h}_{t-1} = \frac{1}{\sqrt{\alpha_t}}(\mathbf{h}_t + (1-\alpha_t)\nabla_{\mathbf{h}_t} \log p_t(\mathbf{h}_t|\mathbf{y})), \tag{6}$$

where $\nabla_{\mathbf{h}_t} \log p_t(\mathbf{h}_t|\mathbf{y})$ is posterior score decomposed by Bayes' rule as,

$$\nabla_{\mathbf{h}_t} \log p_t(\mathbf{h}_t|\mathbf{y}) = \nabla_{\mathbf{h}_t} \log p_t(\mathbf{h}_t) + \nabla_{\mathbf{h}_t} \log p_t(\mathbf{y}|\mathbf{h}_t), \tag{7}$$

Prior score can be approximated by trained denoising networks expressed as,

$$\nabla_{\mathbf{h}_t} \log p_t(\mathbf{h}_t) \approx -\frac{1}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{h}_t, t). \tag{8}$$

Likelihood score can be approximated by Gaussian approximation, expressed as,

$$p(\mathbf{y}|\mathbf{h}_t) = \int p(\mathbf{y}|\mathbf{h})p(\mathbf{h}|\mathbf{h}_t)d\mathbf{h} \approx \mathcal{N}(\mathbf{y}|\mathbb{E}[\mathbf{h}|\mathbf{h}_t], \mathbb{V}[\mathbf{h}|\mathbf{h}_t]) \tag{9}$$

Mean and variance of Gaussian distribution is approximated by the Tweedie's formula [10].

$$\mathbb{E}[\mathbf{h}_0|\mathbf{h}_t] = \frac{1}{\sqrt{\bar{\alpha}_t}}(1 - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\mathbf{h}_t, t)), \tag{10}$$

$$\mathbb{V}[\mathbf{h}_0|\mathbf{h}_t] = \mathbf{\Sigma}_t \nabla_{\mathbf{h}_t}^\top \mathbb{E}[\mathbf{h}_0|\mathbf{h}_t], \tag{11}$$

Likelihood score can be approximated through Gaussian approximation and Tweedie's formula using well-trained score networks.

$$\begin{aligned} &\nabla_{\mathbf{h}_t} \log p_t(\mathbf{y}|\mathbf{h}_t) \\ \approx &(\mathbf{\Sigma}_\mathbf{y} + \mathbf{A}\mathbf{\Sigma}_t^2 \nabla_{\mathbf{h}_t}^\top \mathbb{E}[\mathbf{h}_0|\mathbf{h}_t]\mathbf{A}^\top)^{-1}(\mathbf{y} - \mathbf{A}\mathbb{E}[\mathbf{h}_0|\mathbf{h}_t]) \end{aligned} \tag{12}$$

Due to high complexity of calculation of the inverse matrix, GMRES algorithm is adopted to reduce the complexity [11].

$$(\mathbf{\Sigma}_\mathbf{y} + \mathbf{A}\mathbf{\Sigma}_t^2 \nabla_{\mathbf{h}_t}^\top \mathbb{E}[\mathbf{h}_0|\mathbf{h}_t]\mathbf{A}^\top)\mathbf{u} = \mathbf{y} - \mathbf{A}\mathbb{E}[\mathbf{h}_0|\mathbf{h}_t], \tag{13}$$

where $\mathbf{u}$ is .

IV. Simulations

V. Conclusion

In this letter, the novel channel estimation algorithm has been proposed to handle the accuracy, pilot overhead, and complexity simultaneously.

TABLE I
Hyper-parameter settings for simulation

Parameter	Value
Batch Size	256
Optimizer	AdamW
Learning Rate	0.0001
Decaying Rate	0.1 (per 50 epoch)
Training Epoch	500

TABLE II
Computational complexity for diffusion model-based channel estimators in terms of FLOPs, NFE, and latency

Method	FLOPs	NFE	Latency (s)
SGM [7]	1.028×10^{12}	6933	84.97
DMPS [9]	2.449×10^{10}	100	1.72
Proposed	4.899×10^{9}	20	2.39

References

[1] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[2] P. Zhang, L. Gan, C. Ling, and S. Sun, "Atomic Norm Denoising-Based Joint Channel Estimation and Faulty Antenna Detection for Massive MIMO," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1389–1403, Feb. 2018.

[3] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO Architectures for Millimeter Wave Communications: Phase Shifters or Switches?" *IEEE Access*, vol. 4, pp. 247–267.

[4] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed Sensing for Wireless Communications: Useful Tips and Tricks," *IEEE Commun. Surv. Tutor.*, vol. 19, no. 3, pp. 1527–1550, 2017.

[5] W. Kim, Y. Ahn, J. Kim, and B. Shim, "Towards deep learning-aided wireless channel estimation and channel state information feedback for 6G," *J. Commun. Netw.*, vol. 25, no. 1, pp. 61–75, Feb. 2023.

[6] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep Learning-Based Channel Estimation for Beamspace mmWave Massive MIMO Systems," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 852–855, Oct. 2018.

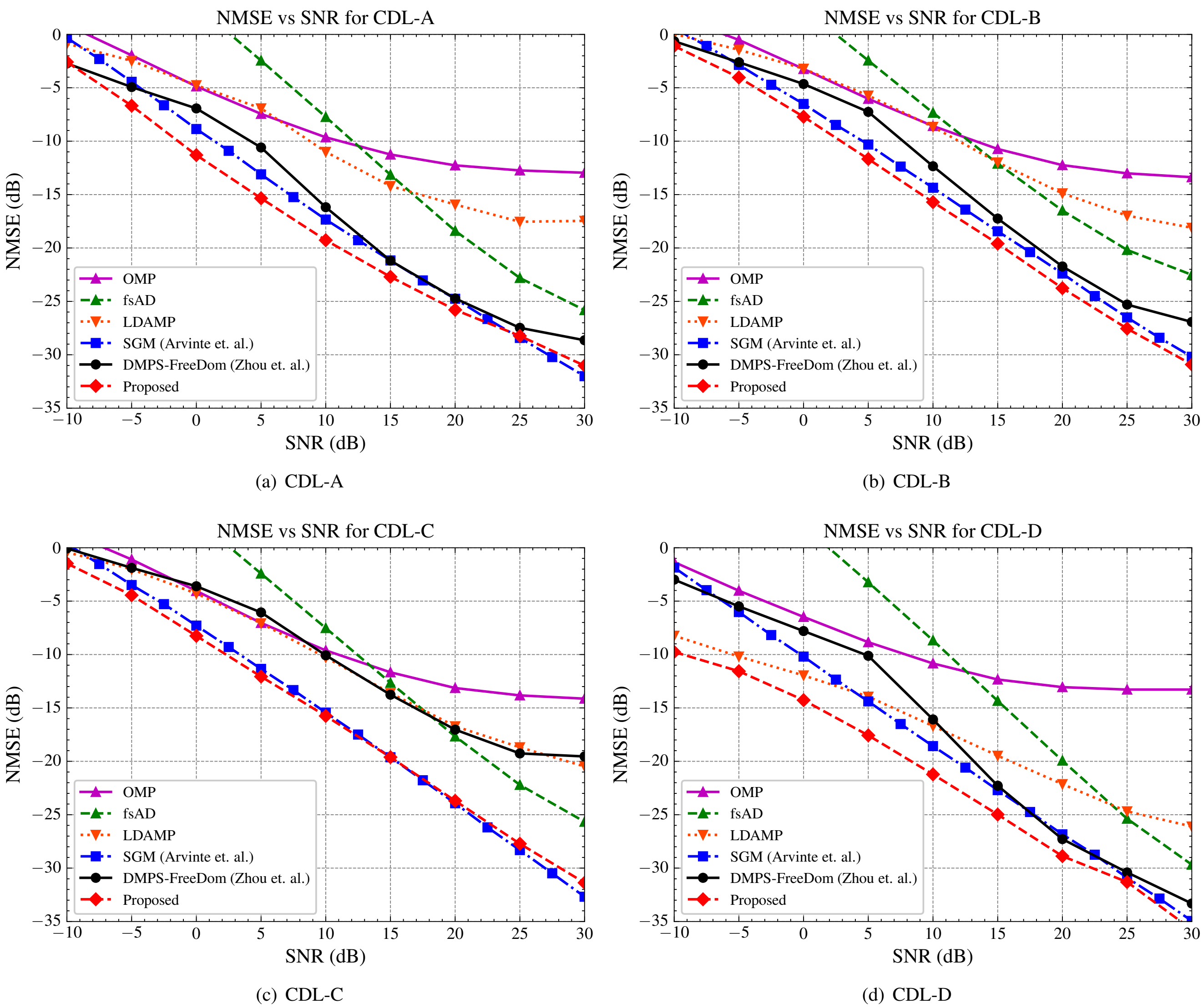
[7] M. Arvinte and J. I. Tamir, "MIMO channel estimation using score-based generative models," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 6, pp. 3698–3713.

[8] B. Fesl, M. Baur, F. Strasser, M. Joham, and W. Utschick, "Diffusion-based generative prior for low-complexity MIMO channel estimation," *IEEE Wirel. Commun. Lett.*, vol. 13, no. 12, pp. 3493–3497.

[9] X. Zhou, L. Liang, J. Zhang, P. Jiang, Y. Li, and S. Jin, "Generative diffusion models for high dimensional channel estimation," *IEEE Trans. Wirel. Commun.*, pp. 1–1.

[10] B. Efron, "Tweedie's formula and selection bias," *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1602–1614, 2011.

[11] Y. Saad and M. H. Schultz, "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems," *SIAM Journal on Scientific and Statistical Computing*, vol. 7, no. 3, pp. 856–869, Jul. 1986.



(a) CDL-A (b) CDL-B (c) CDL-C (d) CDL-D

Fig. 1. Channel estimation performance in terms of NMSE with ρ=0.6 (a) CDL-A. (b) CDL-B. (c) CDL-C. (d) CDL-D.