REGULARIZED NEURAL DETECTION FOR MILLIMETER WAVE MASSIVE MIMO COMMUNICATION SYSTEMS WITH ONE-BIT ADCS

Aditya Sant and Bhaskar D. Rao

Department of Electrical and Computer Engineering, University of California San Diego

ABSTRACT

Multi-user massive MIMO signal detection from one-bit received measurements strongly depends on the wireless channel. To this end, majority of the model and learning-based approaches address detector design for the rich-scattering, homogeneous Rayleigh fading channel. Our work proposes detection for one-bit massive MIMO for the lower diversity mmWave channel. We analyze the limitations of the current state-of-the-art gradient descent (GD)-based joint multi-user detection of one-bit received signals for the mmWave channels. Addressing these, we introduce a new framework to ensure equitable per-user performance, in spite of joint multi-user detection. This is realized by means of: *(i)* a parametric deep learning system, i.e., the mmW-ROBNet, *(ii)* a constellation-aware loss function, and *(iii)* a hierarchical detection training strategy. The experimental results corroborate this proposed approach for equitable per-user detection.

Index Terms— Millimeter Wave Multiple Input Multiple Output, One-bit ADCs, Deep Learning, Algorithm Unrolling

1. INTRODUCTION

Massive MIMO 5G systems, and beyond, propose an entirely new age of communication systems and interconnected devices [1]. Although massive MIMO systems hold immense advantages for highly directional mmWave communication [2, 3], the resulting large number of analog to digital converters (ADCs) required present a significant challenge for system cost and widescale deployment [4, 5]. In order to scale the number of RF chains with massive MIMO, the use of few-bit ADCs presents a promising solution to system cost and implementation. A special case of few-bit ADCs, i.e., the one-bit ADC, has gained increased interest for both channel estimation as well as end-to-end detection. To this end, this work focuses on mmWave MIMO neural detection for one-bit received signals.

Beginning with the linear system characterization through Bussgang's theorem [6], a large number of model-based detectors have been proposed for both single carrier and multi-carrier systems using one-bit ADCs [7–11]. A marked boost in this area was provided via GD-based strategies, beginning with the formulation of the one-bit likelihood [12]. The near maximum likelihood (n-ML) detector efficiently implemented the first such GD-based detector [13]. The work in [14], further improved on the likelihood formulation and proposed a robust learning-based detector, OBMNet, inspiring the new class of learning-based detectors via deep neural networks (DNNs).

DNNs, via a general parametric structure and ability for universal functional approximation [15, 16], provide an encouraging direction for general MIMO receiver design [17, 18] and in particular for detector design for one-bit massive MIMO because of the inherent nonlinearity in the measurement process [19–22]. Our recent work [23] also leveraged this to build a novel DNN-aided regularized GD detector - the ROBNet, as well as a constellation-aware loss function, for one-bit massive MIMO. However, majority of both model and learning-based approaches are applied to the one-bit receiver under the Rayleigh-fading MIMO channel, with very limited work specifically tailored to the mmWave channel [24–26].

Through this work, we empirically analyze the challenges to signal detection from one-bit data for the mmWave MIMO channel. The regularized GD detection framework [23], wherein we utilize a DNN-aided regularization to augment each GD iteration, is appropriately modified to address these challenges. The unfolded DNN implementation, the mmW-ROBNet, capitalizes on the advantages of the regularized GD framework specifically for the mmWave channel. Additionally, we propose a novel hierarchical detection strategy to ensure equitable per-user detection, in spite of the challenges for joint multi-user detection. Our experimental results highlight the utility of the proposed signal detection strategy from one-bit data.

2. SYSTEM MODEL AND GD-BASED DETECTION

This section describes the mmWave channel and one-bit receiver model, followed by an overview of the OBMNet [14], one of the important blocks for our proposed detector using regularized GD.

2.1. System model - MmWave channel and one-bit receiver

We begin by describing the K-user sectored LOS mmWave channel model [3]. The channel model is expressed as

$$\bar{\mathbf{H}} = \left[\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_K)\right] \cdot \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_K), \quad (1)$$

where each $\mathbf{a}(\theta_i)$ is the uniform linear array (ULA) manifold with N antennas for the i^{th} user. Each user path gain α_i and path angle θ_i is independently drawn from the distributions $\mathcal{N}(0,1)$ and $\mathcal{U}(-\pi/3, \pi/3)$, respectively. We order the four users 1 to 4 in decreasing order of received channel powers[1], i.e., $\{|\alpha_i|^2\}_{i=1}^K$. The received signal at the base-station (BS), before quantization, is

$$\bar{\mathbf{r}} = \bar{\mathbf{H}}\bar{\mathbf{x}} + \bar{\mathbf{n}}, \quad (2)$$

where $\bar{\mathbf{x}}$ is the vector of M-QAM constellation symbols and $\bar{\mathbf{n}}$ is the AWGN, whose variance determines the signal-to-noise ratio (SNR).

In order to express the algorithm design as a function of real-valued inputs, we introduce operators $\mathcal{T}_{\mathrm{m}}(\cdot)$ and $\mathcal{T}_{\mathrm{v}}(\cdot)$ to convert matrices and vectors, respectively, into the real valued forms. Using these, we convert the received signal (2) and the observed channel

This work is financially supported in part by ONR Grant No. N00014-18-1-2038, NSF grant CCF-2124929, NSF Grant CCF-2225617, and the UCSD Center for Wireless Communications.

[1] User ordering done during CSI acquisition and initial access

matrix (1) into real-valued forms as [12–14]

$$\mathbf{H} = \mathcal{T}_{\rm m}(\bar{\mathbf{H}}) = \begin{bmatrix} \Re\mathrm{e}(\bar{\mathbf{H}}) & -\Im\mathrm{m}(\bar{\mathbf{H}}) \\ \Im\mathrm{m}(\bar{\mathbf{H}}) & \Re\mathrm{e}(\bar{\mathbf{H}}) \end{bmatrix}, \mathbf{r} = \mathcal{T}_{\rm v}(\bar{\mathbf{r}}) = \begin{bmatrix} \Re\mathrm{e}(\bar{\mathbf{r}}) \\ \Im\mathrm{m}(\bar{\mathbf{r}}) \end{bmatrix},$$
$$\mathbf{x} = \mathcal{T}_{\rm v}(\bar{\mathbf{x}}) = \begin{bmatrix} \Re\mathrm{e}(\bar{\mathbf{x}}) \\ \Im\mathrm{m}(\bar{\mathbf{x}}) \end{bmatrix}, \quad \mathbf{n} = \mathcal{T}_{\rm v}(\bar{\mathbf{n}}) = \begin{bmatrix} \Re\mathrm{e}(\bar{\mathbf{n}}) \\ \Im\mathrm{m}(\bar{\mathbf{n}}) \end{bmatrix}.$$

with $\Re\mathrm{e}(\cdot)$ and $\Im\mathrm{m}(\cdot)$ the real and imaginary parts, respectively. All subsequent likelihood and detection algorithm expressions are henceforth expressed in the real-valued form, using $\mathcal{T}_{\rm m}(\cdot)$ and $\mathcal{T}_{\rm v}(\cdot)$. One-bit quantization at the BS transforms the received signal (2) as

$$\mathbf{y} = \mathrm{sign}(\mathbf{H}\mathbf{x} + \mathbf{n}) = \mathrm{sign}(\mathbf{r}). \tag{3}$$

We assume that the underlying detection algorithm has perfect channel state information $\mathbf{H}$, similar to the benchmark works [13, 14].

2.2. GD-based detection for general one-bit MIMO

The exact one-bit maximum likelihood (ML) optimization is derived in [12], based on the standard normal cumulative distribution function (cdf). However, optimizing this cdf-based cost function runs into instabilities, especially at high SNRs [14, 27]. Using the popular DNN activation, the logistic sigmoid, an approximation of the Gaussian cdf-based ML formulation is derived as [28]

$$\hat{\mathbf{x}}_{\rm ML} = \underset{\mathbf{x}\in\mathcal{M}^{2K}}{\mathrm{argmin}} \sum_{i=1}^{2N} \log\left(1 + e^{-c\sqrt{2\rho}y_i\mathbf{h}_i^{\rm T}\mathbf{x}}\right), \tag{4}$$

where $\mathcal{M}^{2K}$ is the $2K-$dimensional vector of M-QAM symbols (in real-valued form), $c = 1.702$ and SNR $\rho = \frac{\mathbb{E}(||\bar{\mathbf{H}}\bar{\mathbf{x}}||^2)}{\mathbb{E}(||\bar{\mathbf{n}}||^2)}$. Using an unconstrained GD on the likelihood (4) results in the iterative update

$$\begin{aligned} \mathbf{x}^{(t+1)} &= \mathbf{x}^{(t)} - \alpha^{(t)}\nabla_{\mathbf{x}}^{(t)} \\ &= \mathbf{x}^{(t)} + \alpha^{(t)}\mathbf{G}^{\rm T}\sigma(-\mathbf{G}\mathbf{x}^{(t)}), \quad t = 0, \ldots, T-1, \end{aligned} \tag{5}$$

where $\mathbf{G} = \mathrm{diag}(y_1, y_2, \ldots, y_{2N})\,\mathbf{H}$ and $\sigma(\cdot)$ is the logistic sigmoid. The final estimated symbols $\mathbf{x}^{(T)}$ are projected onto the M-QAM constellation as $\mathbf{x}^{(T)} \leftarrow \frac{\sqrt{K}}{||\mathbf{x}^{(T)}||}\mathbf{x}^{(T)}$. The authors in [14] implement the GD algorithm (5) as a learnable unfolded DNN, i.e., the OBMNet, with step sizes at each iteration $\{\alpha^{(t)}\}_{t=0}^{T-1}$ as the only learnable parameters. The network is trained on the MSE loss $\mathcal{L} = \frac{1}{N_{\rm train}}\sum_{n=1}^{N_{\rm train}} ||\mathbf{x}_n^{(T)} - \tilde{\mathbf{x}}_{{\rm train},n}||^2$.

This framework is extremely robust for lower order M-QAM constellations, i.e., QPSK. However, our related work in [27] [23] presents the limitations of the OBMNet for higher-order M-QAM symbol recovery, for the Rayleigh channel, namely the limited network expressivity and the large recovered constellation cluster spread. We now analyze this approach for the mmWave channel.

3. DNN-AIDED GD FOR MMWAVE ONE-BIT RECEIVERS

This section presents the GD-based detection tailored specifically to the mmWave channel. Beginning with the challenges for joint detection, we describe our proposed approach and implementation.

3.1. Challenges to joint detection for the mmWave channel

Different from the rich scattering of the Rayleigh-fading channel, there is lower diversity in the mmWave channel due to antenna correlation [2, 3]. This is pictorially represented in Fig. 1 by the histograms of the square root of channel power (see eq. (1)), i.e., $||\bar{\mathbf{h}}_i||$, in decreasing order from User 1 ($||\bar{\mathbf{h}}_1||$) to User 4 ($||\bar{\mathbf{h}}_4||$). These plots portray a larger spread in mmWave channel powers per user, vs the more equitable power distribution of the Rayleigh channel.

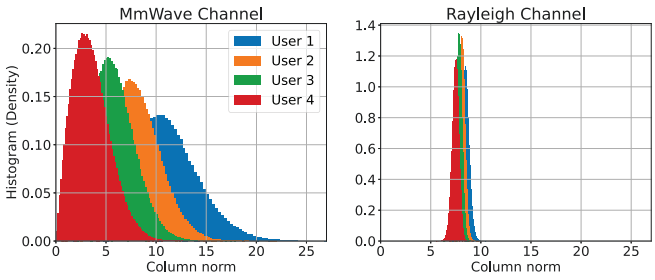

Fig. 1. Distribution of the square root power for mmWave (left) and Rayleigh-fading channel (right) with $N = 64$ antennas, $K = 4$ users. Here User 1 has the strongest channel and User 4 the weakest.

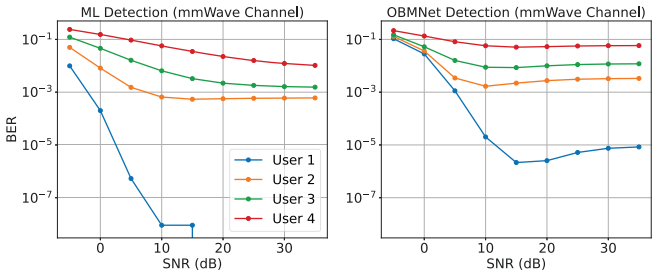

Fig. 2. Comparing performance of ML (left) & OBMNet [14] (right) detection for mmWave channel (1) with $K = 4$ users, $N = 64$ antennas, each user transmitting QPSK symbols. User 1 has the strongest channel and User 4 the weakest.

The impact of this channel power spread among the users affects detection performance, i.e., bit error rate (BER), of each user. This is quantitatively illustrated in Fig. 2, comparing maximum likelihood (ML) detection, using exhaustive search in $\mathcal{M}^{2K}$, to the OBMNet [14], for QPSK symbols. The left plot in Fig. 2 shows the overt disparity in the ML detection performance for the different users; User 1, with the strongest channel, experiences more than six orders of magnitude lower BER, at saturation, compared to User 4, with the weakest channel. The right plot in Fig. 2 highlights that the OBMNet [14], optimizing for the joint detection performance, is limited by the weakest users, thus unable to equitably handle users with stronger channels. The next sub-sections describe the design of a DNN-aided detector to provide a more equitable BER performance.

3.2. Robust detection framework for mmWave channels

Signal detection for the receiver (1)-(3) is approached via regularized DNN-aided GD, with a novel hierarchical detection training strategy. Our approach consists of three major components, elaborated here.

3.2.1. User-matched regularized GD

The regularized GD detection for one-bit receivers [23] augments the GD step (5) via a DNN-aided projection step, thereby fine-tuning each iteration. For mmWave channels, we devise a modification - a user-matched regularized GD. Herein, we incorporate the mmWave

channel path gains per-user $\{\alpha_i\}_{i=1}^K$, from (1), as

$$\hat{\nabla}_{\mathbf{x}}^{(t)} = \mathcal{T}_{\mathrm{m}}\big(\mathrm{abs}(\mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_K)^{-1})\big)\, \nabla_{\mathbf{x}}^{(t)} \tag{6a}$$

$$\hat{\mathbf{x}}^{(t+1)} = \mathbf{x}^{(t)} - \alpha^{(t)} \hat{\nabla}_{\mathbf{x}}^{(t)} \tag{6b}$$

$$\mathbf{x}^{(t+1)} = \hat{\mathbf{x}}^{(t+1)} + h_{\phi}^{(t)}(\mathbf{x}^{(t)}, \hat{\nabla}_{\mathbf{x}}^{(t)}, \hat{\mathbf{x}}^{(t+1)}). \tag{6c}$$

The user-matching step (6a) homogenizes the gradient from (5), stabilizing the unequal channel power scaling, among the users [2]. Update steps (6b)-(6c) execute regularized GD with the user-matched gradient. The function $h_{\phi}^{(t)}$ is the DNN-based parametric regularization at iteration t, elaborated further in Sec. 3.3. The regularized GD algorithm (6), implemented as an unfolded DNN, i.e., the mmWave regularized one-bit net (mmW-ROBNet), is illustrated in Fig. 3.

3.2.2. *Constellation-aware DNN loss function*

In order to tailor the DNN loss function to the M-QAM symbol recovery, we incorporate the regularized loss function [23]

$$\mathcal{L} = \frac{1}{N_{\mathrm{train}}} \sum_{n=1}^{N_{\mathrm{train}}} \big[||\mathbf{x}_n^{(T)} - \tilde{\mathbf{x}}_{\mathrm{train},n}||^2 + \lambda\, \mathcal{R}(\mathbf{x}_n^{(T)}, \tilde{\mathbf{x}}_{\mathrm{train},n})\big], \tag{7}$$

where $\mathcal{R}(\cdot)$ is a constellation-aware regularization, based on a smooth quantization of the network output, $\mathcal{Q}_{\beta}(\cdot)$, implemented as

$$\mathcal{R}(\mathbf{x}_n^{(T)}, \tilde{\mathbf{x}}_{\mathrm{train},n}) = ||\mathcal{Q}_{\beta}(\mathbf{x}_n^{(T)}) - \tilde{\mathbf{x}}_{\mathrm{train},n}||^2. \tag{8}$$

For QPSK symbol recovery[3], this is given by $\mathcal{Q}_{\beta}(x) = \tanh(\beta\, x)$. This saturating nonlinearity $\mathcal{Q}_{\beta}(\cdot)$ attenuates symbol errors within the true constellation boundaries and amplifies errors for crossing over symbol boundaries. The constellation-aware regularization (8) thus incorporates symbol error rate (SER), in addition to MSE, into the total loss (7), resulting in a robust training.

3.2.3. *Hierarchical detection training*

The final block for robust DNN-aided GD is the enhanced DNN training procedure. In particular, we control the trajectory of intermediate GD iterates $\{\mathbf{x}^{(t)}\}_{t=1}^{T-1}$ in (6), essential to the final estimate $\mathbf{x}^{(T)}$. To this end, we propose a sub-loss $\mathcal{L}^{(t)}$ at each iteration t as

$$\mathcal{L}^{(t)} = \frac{1}{N_{\mathrm{train}}} \sum_{n=1}^{N_{\mathrm{train}}} \big[||\mathbf{w}^{(t)} \odot \mathbf{x}_n^{(t)} - \mathbf{w}^{(t)} \odot \tilde{\mathbf{x}}_{\mathrm{train},n}||^2 + \lambda\, \mathcal{R}(\mathbf{w}^{(t)} \odot \mathbf{x}_n^{(t)}, \mathbf{w}^{(t)} \odot \tilde{\mathbf{x}}_{\mathrm{train},n})\big] \tag{9}$$

where,

$$\mathbf{w}^{(t)} = \tilde{c}^{(t)}\, \big[1, \exp(-\kappa_t), \ldots, \exp(-(K-1)\,\kappa_t)\big]^{\mathrm{T}}.$$

Here $\mathbf{w}^{(t)}$ is the masking vector, κ_t is the user masking coefficient and $\tilde{c}^{(t)}$ is the normalization constant such that $||\mathbf{w}^{(t)}|| = 1$, at the t^{th} iteration. The operator $\odot$ denotes element-wise product. The regularization $\mathcal{R}(\cdot)$ is the same as (8). The total DNN training loss function is given by

$$\mathcal{L} = \sum_{t=1}^{T} \mathcal{L}^{(t)}. \tag{10}$$

[2]Due to space constraints, we omit the analytical proof, but will include it in the complete future publication of this work

[3]Our work in [23] also extends this to 16-QAM. The complete version of this work, will incorporate 16-QAM for mmWave channels as well.

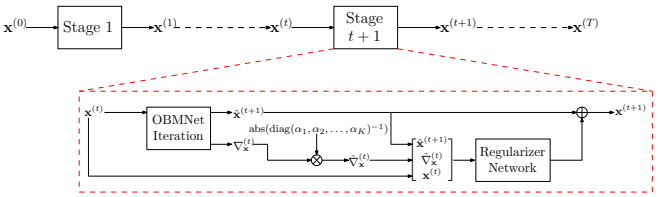

Fig. 3. Block diagram for the mmW-ROBNet

Table 1. DNN Parameters of the Regularizer Network (For K Users)

Network	Layer	Parameters
mmW-ROBNet (each stage)	*Convolution (1-D)*	conv + ReLU + bn
		Input dim - K
		Input chan - 6
		Output chan - 64
		Kernel size - 3
	Fully-connected	Input dim - 64K
		Output dim - 2K
		Hidden layers - 3
		Hidden dim - {128,64,32}
		bn dim - {128,64,32}
		Nonlinearity - ReLU

Owing to the exponential decay over the user index, the masking vector $\mathbf{w}^{(t)}$ attenuates the users that experience a weaker channel, reducing their contribution to the loss function (9). Further, we begin with a large value for the masking coefficient κ_t and decrease this over the GD iterations to $\kappa_T = 0$. This results in the first few regularization sub-networks $h_{\phi}^{(t)}$ being trained to detect the stronger users only. The subsequent sub-networks gradually add users in decreasing order of their channel quality, by jointly detecting these with the stronger users. Finally, the last sub-network $h_{\phi}^{(T)}$ is trained to jointly detecting all the users. Since the training strategy (9)-(10) detects multiple users from the strongest to the weakest, it is called hierarchical detection or *HieDet training*.

3.3. DNN implementation for user-matched regularized GD

Inspired by the potential of model-based algorithm unrolling for general signal processing [14, 21, 29, 30], the unfolded DNN, ROBNet, is developed as an approach to solve the constrained optimization (4) for any general channel matrix $\mathbf{H}$, via a DNN-augmented GD algorithm [23]. The mmW-ROBNet framework (6), introduced in Sec. 3.2.1, incorporates the specific properties of the mmWave channel. The implementation details are elaborated as follows.

(i) The T-stage regularized GD algorithm (6), is unfolded into T distinct sub-networks (each represented as Stage t in Fig. 3).

(ii) At the beginning of each Stage $t+1$, the OBMNet iteration (5) generates the t^{th} gradient and output $\nabla_{\mathbf{x}}^{(t)}$ and $\hat{\mathbf{x}}^{(t+1)}$, respectively.

(iii) The mmWave-channel powers per user, $\{|\alpha_i|^2\}_{i=1}^K$, user-matches the OBMNet-generated gradient $\nabla_{\mathbf{x}}^{(t)}$ as (6a), to get $\hat{\nabla}_{\mathbf{x}}^{(t)}$

(iv) The previous estimate $\mathbf{x}^{(t)}$, user-matched gradient $\hat{\nabla}_{\mathbf{x}}^{(t)}$ and OBMNet output $\hat{\mathbf{x}}^{(t+1)}$ is passed to the Regularization Network $h_{\phi}^{(t)}$ for fine-tuning (see Table 1 for DNN parameters).

(v) Additionally, a residual link from the OBMNet output is fed to the output of the Regularization Network, thereby imparting a stage-dependent correction, to the unconstrained OBMNet step.

(vi) The final output $\mathbf{x}^{(T)}$ is normalized as $\mathbf{x}^{(T)} \leftarrow \frac{\sqrt{2K}}{||\mathbf{x}^{(T)}||}\mathbf{x}^{(T)}$.

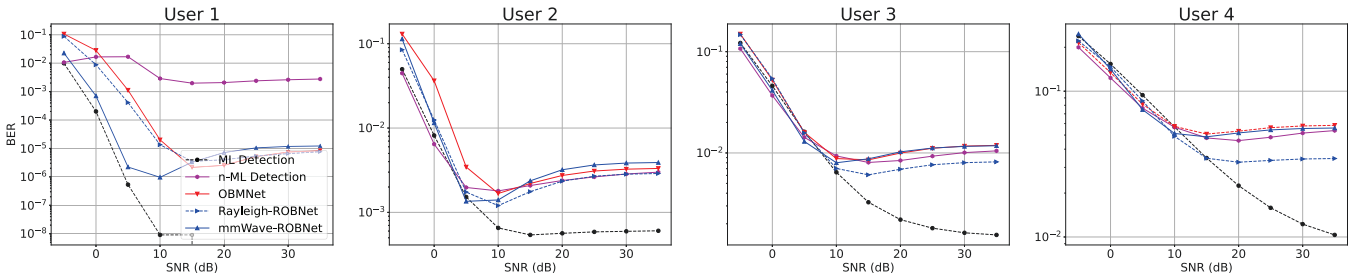

Fig. 4. Comparison of per-user BER Vs SNR performance, from the strongest user, User 1, to the weakest, User 4. All $K = 4$ users, transmitting QPSK symbols, are jointly detected at the BS with $N = 64$ antennas

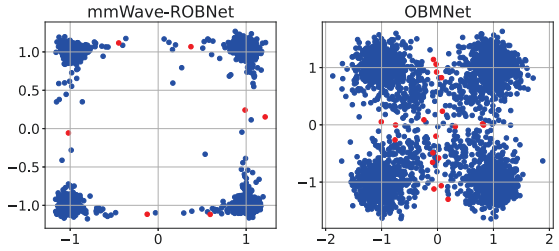

Fig. 5. Recovered constellation for joint detection of all $K = 4$ users, received at ULA with $N = 64$ antennas at $\mathrm{SNR} = 15\,\mathrm{dB}$

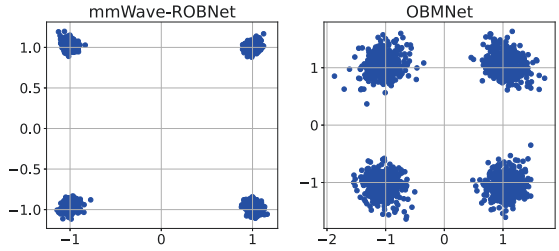

Fig. 6. Recovered constellation for User 1 and User 2 only, received at ULA with $N = 64$ antennas at $\mathrm{SNR} = 15\,\mathrm{dB}$

4. EXPERIMENTAL RESULTS

Simulation setup: The sectored mmWave channel [3] is considered, as described in Sec. 2.1. The signal from $K = 4$ users, transmitting QPSK symbols, is received at a BS ULA with $N = 64$ antennas.
DNN hyperparameters and training: We follow the same DNN training procedure, namely the minibatch description, optimizer and weight decay, as described in our previous work [23]. The mmW-ROBNet consists of $T = 5$ iterations. For the constellation-aware loss function (7)-(8), we assign $\lambda = 5$ and $\beta = 2$. All networks are trained at an intermediate SNR of 15 dB. Finally, the sequence of user masking coefficients $\{\kappa_t\}_{t=1}^{5}$, in (9), is given by $\{10, 1, 0.5, 0.1, 0\}$.
Performance benchmarks: The OBMNet [14], with $T = 20$ iterations is considered as our primary benchmark. We also compare our framework against the the n-ML [13] algorithm optimizing the cdf likelihood. The Rayliegh-ROBNet framework, from our previous work [23], is also considered. Finally, the ML detection, based on an exhaustive search, forms the lower limit of the BER performance.

4.1. Recovered constellation: Scatterplots

We begin with a qualitative performance comparison of the proposed mmW-ROBNet and the benchmark unregularized GD, i.e., the OBMNet [14]. The recovered constellation via joint detection of all four users is provided in Fig. 5 (red dots represent incorrectly detected symbols). Based on these plots, it is evident that both the mmW-ROBNet as well as the OBMNet have poor recovery with bit errors for joint four-user detection. Since all the users experience different quality channels, the recovered constellation for the weaker users is evidently responsible for this performance limitation. Reinforcing this rationale, we also illustrate the scatterplots of the strongest two users, i.e., User 1 and User 2, in Fig. 6. Due to the better quality channel for these users, the mmW-ROBNet, with HieDet training, is able to generate more uniform constellation clusters with markedly reduced cluster spread compared to the OBMNet.

4.2. Detection performance for general mmWave channel

The per-user BER vs SNR performance is given in Fig. 4. As seen from these plots, the sigmoid-based likelihood (4) improves on the cdf-based n-ML detector. Although our Rayleigh-ROBNet [23] outperforms the benchmark detectors, it is also unable to captalize on the channel with the highest power, i.e., User 1. Note that the BER average across the users will not capture this difference in detector performance. The combination of the user-matched GD, along with the sequential HieDet training is able to capitalize on the users with a better channel without significantly affecting users with worse channel quality. Finally, all approaches saturate in BER for high SNR, and reducing the gap to ML is still part of our ongoing research.

5. CONCLUSIONS

Through this work we have presented a novel one-bit neural detection approach, mmW-ROBNet, specifically tailored to the mmWave channel model. We have illustrated the dependence of the detection performance, per user, on the channel power for that user. Addressing this limitation, we have modified the existing regularized GD framework, the ROBNet, to overcome these challenges. In particular, by means of the user-matched regularized GD and HieDet training, we are able to capitalize on the stronger user channel powers for an equitable performance among the multiple users.

Future work in this domain involves improving on the robustness of the HieDet training to further close the gap to ML detection. In addition, we also envision extending this detection to higher order M-QAM constellations, addressing the challenges therin.

6. REFERENCES

[1] K. Shafique, B. A. Khawaja, F. Sabir, S. Qazi, and M. Mustaqim, "Internet of things (iot) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5g-iot scenarios," *Ieee Access*, vol. 8, pp. 23 022–23 040, 2020.

[2] E. Björnson, J. Hoydis, L. Sanguinetti *et al.*, "Massive mimo networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[3] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[4] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath, "Hybrid architectures with few-bit adc receivers: Achievable rates and energy-rate tradeoffs," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2274–2287, 2017.

[5] B. Murmann *et al.*, "Adc performance survey 1997-2020," in *IEEE Int. Solid-State Circuits Conf.(ISSCC) Dig. Tech. Papers VLSI Symp*, 2020.

[6] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted gaussian signals," 1952.

[7] A. Mezghani, M.-S. Khoufi, and J. A. Nossek, "A modified mmse receiver for quantized mimo systems," *Proc. ITG/IEEE WSA, Vienna, Austria*, pp. 1–5, 2007.

[8] Q. Wan, J. Fang, H. Duan, Z. Chen, and H. Li, "Generalized bussgang lmmse channel estimation for one-bit massive mimo systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 4234–4246, 2020.

[9] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive mimo systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 4075–4089, 2017.

[10] L. V. Nguyen, A. L. Swindlehurst, and D. H. Nguyen, "Svm-based channel estimation and data detection for one-bit massive mimo systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2086–2099, 2021.

[11] S. S. Thoota and C. R. Murthy, "Variational bayes' joint channel estimation and soft symbol decoding for uplink massive mimo systems with low resolution adcs," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3467–3481, 2021.

[12] J. Choi, D. J. Love, D. R. Brown, and M. Boutin, "Quantized distributed reception for mimo wireless systems using spatial multiplexing," *IEEE Transactions on Signal Processing*, vol. 63, no. 13, pp. 3537–3548, 2015.

[13] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive mimo systems with one-bit adcs," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 2005–2018, 2016.

[14] L. V. Nguyen, A. L. Swindlehurst, and D. H. Nguyen, "Linear and deep neural network-based receivers for massive mimo systems with one-bit adcs," *IEEE Transactions on Wireless Communications*, vol. 20, no. 11, pp. 7333–7345, 2021.

[15] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[17] M. Khani, M. Alizadeh, J. Hoydis, and P. Fleming, "Adaptive neural signal detection for massive mimo," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5635–5648, 2020.

[18] K. Pratik, B. D. Rao, and M. Welling, "Re-mimo: Recurrent and permutation equivariant neural mimo detection," *IEEE Transactions on Signal Processing*, vol. 69, pp. 459–473, 2020.

[19] S. Kim, M. So, N. Lee, and S. Hong, "Semi-supervised learning detector for mu-mimo systems with one-bit adcs," in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2019, pp. 1–6.

[20] L. V. Nguyen, D. H. Nguyen, and A. L. Swindlehurst, "Dnn-based detectors for massive mimo systems with low-resolution adcs," in *ICC 2021-IEEE International Conference on Communications*. IEEE, 2021, pp. 1–6.

[21] S. Khobahi, N. Shlezinger, M. Soltanalian, and Y. C. Eldar, "Lord-net: Unfolded deep detection network with low-resolution receivers," *IEEE Transactions on Signal Processing*, vol. 69, pp. 5651–5664, 2021.

[22] S. Kim, J. Chae, and S.-N. Hong, "Machine learning detectors for mu-mimo systems with one-bit adcs," *IEEE Access*, vol. 8, pp. 86 608–86 616, 2020.

[23] A. Sant and B. D. Rao, "Regularized neural detection for one-bit massive mimo communication systems," *IEEE Transactions on Machine Learning in Communications and Networking*, (submitted to journal and available on Arxiv).

[24] C. Stöckle, J. Munir, A. Mezghani, and J. A. Nossek, "Channel estimation in massive mimo systems using 1-bit quantization," in *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2016, pp. 1–6.

[25] C.-L. Liu and P. Vaidyanathan, "One-bit sparse array doa estimation," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 3126–3130.

[26] A. Sant and B. D. Rao, "Doa estimation in systems with nonlinearities for mmwave communications," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 4537–4541.

[27] D. Ho, "Channel estimation and data detection methods for 1-bit massive mimo systems," Ph.D. dissertation, University of California San Diego, 2022.

[28] S. R. Bowling, M. T. Khasawneh, S. Kaewkuekool, and B. R. Cho, "A logistic approximation to the cumulative normal distribution," *Journal of Industrial Engineering and Management*, vol. 2, no. 1, pp. 114–127, 2009.

[29] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021.

[30] A. Sant, A. Abdi, and J. Soriaga, "Deep sequential beamformer learning for multipath channels in mmwave communication systems," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 5198–5202.