Learning-Based Signal Detection for Wireless OAM-MIMO Systems With Uniform Circular Array Antennas

NORIFUMI KAMIYA[ID], (Member, IEEE)
System Platform Research Laboratories, NEC Corporation, Kawasaki 211-8666, Japan

e-mail: nori_kamiya@nec.com

ABSTRACT This paper presents a neural-like network-based signal detection method for orbital angular momentum multiplexing systems with uniform circular array antennas. The signal detection network is derived by unfolding the alternating direction method of multipliers (ADMM), and in addition, a parallel interference cancellation (PIC) function is integrated, which enhances the tolerance to inter-mode interference while keeping the complexity feasible. The number of parameters to be learned in each layer of the network is a linear order of the number of antenna elements. Simulation results show that the ADMM-PIC detector exhibits excellent error performance, which cannot be achieved by a conventional minimum mean square error-based detector.

INDEX TERMS Interference cancellation, millimeter wave communication, MIMO, neural network learning, orbital angular momentum (OAM), uniform circular array (UCA).

I. INTRODUCTION

Improving data transmission capacity with multi-input and multi-output (MIMO) multiplexing technologies [1] has become increasingly important for wireless communication systems to meet continuous demand for higher data rates over limited frequency resources. Recently, MIMO multiplexing using orbital angular momentum (OAM) [2], [3] has attracted attention for application to high capacity wireless communication systems, especially to point-to-point (PTP) line-of-sight (LOS) micro-wave/mm-wave radio systems for mobile fronthaul and backhaul links [4]–[7].

The OAM-MIMO multiplexing is realized by utilizing electro-magnetic waves with different OAM-modes, which are inherently orthogonal to each other. Several methods have been proposed to generate OAM waves, such as helicoidal parabolic antennas [3], spiral phase plates [8]–[10] and thin metamaterial plates [11]. Among them, uniform circular array antenna (UCA) is considered as one of most promising because of its simple structure and low cost. Indeed, several experimental studies have demonstrated the effectiveness and feasibility of OAM-MIMO multiplexing systems using UCAs [4]–[7], [12]. Moreover, [13] has investigated

theoretically the UCA-based OAM-MIMO channel and has shown an equivalence to a precoded MIMO channel in terms of channel capacity.

In this paper, we consider OAM-MIMO multiplexing systems with UCAs (especially with multi-ring UCAs [4]–[6], [14]) and investigate its channel capacity as a precoded MIMO system. The set of OAM-modes for a multi-ring UCA-based OAM-MIMO channel can be thought of as a collection of LOS-MIMO channels [15]. While these LOS-MIMO channels are ideally orthogonal to each other, component subchannels in each LOS-MIMO channel are not orthogonally multiplexed in general. Thus, conventional MIMO receiver with mode-wise zero-forcing (ZF) or minimum mean square error (MMSE) detector will not give optimum performance for the OAM-MIMO systems. Furthermore, the mode-wise capacity as well as signal detector performance deteriorates seriously when transmitter (Tx) and receiver (Rx) UCAs are not ideally aligned. The antenna misalignment is likely to be time-varying due to various reasons like when they are mounted on a tower that sways in the wind, etc. It is thus of great concern to alleviate the performance degradation due to the misalignment, and this has been studied in [16]–[18] for the single-ring UCA case. To address the above issue, in this paper we consider to apply machine learning techniques.

The associate editor coordinating the review of this manuscript and approving it for publication was Jinming Wen[ID].

Machine learning techniques such as deep learning have recently been widely and successfully applied to many research areas in communication systems, and they have also been being applied to solve research problems in physical-layer communications. In [19], [20], the MIMO detection problem has been investigated in a machine learning framework, and a deep learning network, referred to as the DetNet, has been presented for MIMO detection. The DetNet consists of multiple layers, each of which corresponds to an iteration of a projected gradient descent algorithm. While it has been demonstrated that the DetNet performs well for signal detection of MIMO channels, it has a large number of parameters needed to be trained, which incurs a high computational cost for training. In [21], a different network architecture has been presented, which is derived by unfolding the Alternating Direction Method of Multipliers (ADMM) iterative algorithm [22], [23]. It has been shown by simulations that for MIMO channels with BPSK and QPSK modulation, the ADMM-Net in [21] can achieve competitive performance with a small number of trainable parameters.

Inspired by the work in [19]–[21], we present a learning, neural-like network-based signal detection method for UCA-based OAM-MIMO transmission systems. The OAM-MIMO channels discussed in this paper are assumed to be LOS while most of the prior work on machine learning-based MIMO detection has focused on MIMO channels which are assumed to contain rich multipath components but no strong LOS component, and which are characterized by Gaussian random matrices. There are few attempts to apply machine learning techniques to signal detection for MIMO channels where LOS components are dominant. The main contributions of this paper can be summarized as follows: (i) the channel capacity is investigated for multi-ring UCA-based OAM-MIMO multiplexing with mode-wise MMSE detectors. It is shown that the performance of the MMSE detector is sometimes far from optimum and is very sensitive to antenna misalignment. (ii) In order to improve the detector performance, a neural-like network-based signal detector is presented for the OAM-MIMO systems, which is derived by unfolding the ADMM [22]. In addition, in order to enhance the tolerance to antenna misalignment, parallel interference canceller (PIC) is integrated into the unfolded ADMM network. The main difference of the ADMM part of our ADMM-PIC from the ADMM-Net presented in [21] is the selection of the trainable parameters, which contributes to improve the detector performance. It is shown by simulations that the ADMM-PIC detector exhibits excellent error performance even when there is severe inter-mode interference due to antenna misalignment, which cannot be achieved by the conventional MMSE-based detector.

The rest of this paper is organized as follows: In Section II, we briefly review the UCA-based OAM-MIMO transmission model and then investigate its Shannon capacity and mode-wise MMSE capacity. Section III presents a neural-like network-based OAM-MIMO detection algorithm. Section IV presents numerical simulation results for

256 QAM OAM-MIMO channels with single, double and triple-ring UCAs. Section V summarizes our results and concludes the paper.

Notation: Uppercase and lowercase boldface letters denote matrices and (column) vectors, respectively. The superscripts T, $*$, and $\dagger$ stand, respectively, for transposition, element-wise conjugation and Hermitian transpose; and $\Re$ and $\Im$ denote the real and imaginary part, respectively. Additionally, $\boldsymbol{I}$ denotes the identity matrix of an appropriate size, and $\circ$ and $\otimes$ are the Hadamard and Kronecker products, respectively. For a vector $\boldsymbol{x}$, diag($\boldsymbol{x}$) denotes the diagonal matrix with diagonal entries from $\boldsymbol{x}$, and $\|\boldsymbol{x}\|_2$ denotes the ℓ^2-norm of $\boldsymbol{x}$. For a square matrix $\boldsymbol{A}$, tr($\boldsymbol{A}$) denotes the trace of $\boldsymbol{A}$. Finally, we denote the N-point discrete Fourier transform (N-DFT) matrix by $\boldsymbol{F}_N$, i.e., $\boldsymbol{F}_N = (1/\sqrt{N})[\exp(-j2\pi kl/N)]_{0\leq k,l<N}$.

II. SYSTEM MODEL AND CAPACITY
A. OAM-MIMO TRANSMISSION MODEL
We consider an OAM-MIMO multiplexing system equipped with UCAs at Tx and Rx sides. The Tx and Rx UCAs are composed of M concentric circular array antennas, each of which consists of N antenna elements equidistantly arranged on a ring, and we refer it to as (N, M)-UCA. Fig. 1 illustrates an OAM-MIMO system with $(8, 2)$-UCAs. Hereafter, indices i and j denote integers in $\{0, 1, \ldots, M - 1\}$, and similarly, indices k and l denote integers in $\{0, 1, \ldots, N - 1\}$. The (N, M)-UCA transmitter generates and transmits signals of the form of

$$s_i^{(k)} = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} x_i^{(l)} \exp\left(j\frac{2\pi kl}{N}\right), \qquad (1)$$

where $x_i^{(l)}$ denotes the original complex information signal transmitted through OAM mode-l channel and $s_i^{(k)}$ denotes the signal transmitted from the k-th antenna element in the i-th of M rings. Denoting $\boldsymbol{s}_i = (s_i^{(0)}, s_i^{(1)}, \ldots, s_i^{(N-1)})^T$ and $\boldsymbol{x}_i = (x_i^{(0)}, x_i^{(1)}, \ldots, x_i^{(N-1)})^T$, (1) can be written in matrix form as $\boldsymbol{s}_i = \boldsymbol{F}_N^{\dagger}\boldsymbol{x}_i$.
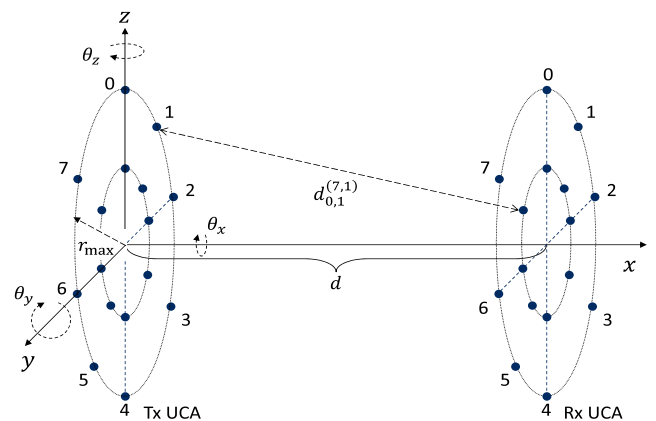

FIGURE 1. OAM-MIMO multiplexing with (8, 2)-UCAs.

The free space transmission model [13], [16]–[18] is used to simulate an OAM-MIMO channel, in which the transfer function between a pair of Tx and Rx antenna elements is written as

$$h(d) = G\frac{\lambda}{4\pi d}\exp\left(-j\frac{2\pi d}{\lambda}\right), \qquad (2)$$

where d is transmission distance, λ is wavelength, $\lambda/(4\pi d)$ represents the free space loss, and G contains all relevant constants such as attenuation and phase rotation caused by antennas and their patterns on both sides. We denote by $d_{i,j}^{(k,l)}$ the distance between the k-th antenna element in the i-th ring on the Rx side and the l-th antenna element in the j-th ring on the Tx side. The channel matrix of an OAM-MIMO system with the (N, M)-UCAs can then be expressed as the following block matrix:

$$\begin{bmatrix} \boldsymbol{H}_{0,0} & \boldsymbol{H}_{0,1} & \cdots & \boldsymbol{H}_{0,M-1} \\ \boldsymbol{H}_{1,0} & \boldsymbol{H}_{1,1} & \cdots & \boldsymbol{H}_{1,M-1} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{H}_{M-1,0} & \boldsymbol{H}_{M-1,1} & \cdots & \boldsymbol{H}_{M-1,M-1} \end{bmatrix}, \qquad (3)$$

where $\boldsymbol{H}_{i,j}$ is an $N\times N$ matrix whose (k, l)-th entry is $h(d_{i,j}^{(k,l)})$, i.e., $\boldsymbol{H}_{i,j} = [h(d_{i,j}^{(k,l)})]_{0\leq k,l<N}$. As illustrated in Fig. 1, we denote by θ_x, θ_y and θ_z, respectively, the misalignment angles around x-, y- and z-axes between Tx and Rx UCAs. Note here that when Tx and Rx UCAs are ideally aligned, i.e., $\theta_x = \theta_y = \theta_z = 0$, $\boldsymbol{H}_{i,j}$'s can be seen as circulant matrices, and thus in this case, $\boldsymbol{F}_N\boldsymbol{H}_{i,j}\boldsymbol{F}_N^\dagger$ becomes a diagonal matrix. We denote by $h_{i,j}^{(k,l)}$ the (k, l)-th entry of the matrix $\boldsymbol{F}_N\boldsymbol{H}_{i,j}\boldsymbol{F}_N^\dagger$.

The signal received at Rx UCA is then post-processed by applying the N-DFT, and thus the (N, M)-UCA receiver generates signals of the form of

$$y_i^{(l)} = \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} r_i^{(k)}\exp\left(-j\frac{2\pi kl}{N}\right), \qquad (4)$$

where $r_i^{(k)}$ denotes the signal received at the k-th antenna element in the i-th ring. Consequently, the OAM-MIMO multiplexing system with (N, M)-UCAs at Tx and Rx sides can be modeled as follows:

$$\begin{bmatrix} \boldsymbol{y}^{(0)} \\ \boldsymbol{y}^{(1)} \\ \vdots \\ \boldsymbol{y}^{(N-1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}^{(0,0)} & \cdots & \boldsymbol{H}^{(0,N-1)} \\ \boldsymbol{H}^{(1,0)} & \cdots & \boldsymbol{H}^{(1,N-1)} \\ \vdots & & \vdots \\ \boldsymbol{H}^{(N-1,0)} & \cdots & \boldsymbol{H}^{(N-1,N-1)} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^{(0)} \\ \boldsymbol{x}^{(1)} \\ \vdots \\ \boldsymbol{x}^{(N-1)} \end{bmatrix} + \boldsymbol{n} \qquad (5)$$

where $\boldsymbol{y}^{(k)}$, $\boldsymbol{x}^{(k)}$ and $\boldsymbol{H}^{(k,l)}$ denote $(y_0^{(k)}, y_1^{(k)}, \ldots, y_{M-1}^{(k)})^T$, $(x_0^{(k)}, x_1^{(k)}, \ldots, x_{M-1}^{(k)})^T$ and $[h_{i,j}^{(k,l)}]_{0\leq i,j<M}$, respectively, and $\boldsymbol{n}$ is a complex Gaussian vector noise $\boldsymbol{n} \sim \mathcal{CN}(0, \sigma^2\boldsymbol{I})$. In the following, we write (5) as $\boldsymbol{y} = \boldsymbol{Hx} + \boldsymbol{n}$ for simplicity. Note that the matrix $\boldsymbol{H}$ can be obtained by multiplying (3) with $\boldsymbol{I}\otimes\boldsymbol{F}_N$ from the left and with $\boldsymbol{I}\otimes\boldsymbol{F}_N^\dagger$ from the right and then by rearranging the columns and rows appropriately. When Tx

and Rx UCAs are ideally aligned, $\boldsymbol{H}$ becomes block diagonal, i.e., $\boldsymbol{H}^{(k,l)} = \boldsymbol{0}$ if $k \neq l$.

It is well-known that the information signal $\boldsymbol{x}$ can be extracted by multiplying the inverse of $\boldsymbol{H}$ by $\boldsymbol{y}$, i.e., $\widehat{\boldsymbol{x}}_{\text{ZF}} = \boldsymbol{H}^{-1}\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{H}^{-1}\boldsymbol{n}$, which is referred to as the ZF detector. An alternative approach is based on MMSE estimation. In an MMSE detector, an estimate of $\boldsymbol{x}$ is given by $\widehat{\boldsymbol{x}}_{\text{MMSE}} = \boldsymbol{Wy}$, where $\boldsymbol{W} = \boldsymbol{PH}^\dagger(\boldsymbol{HPH}^\dagger + \sigma^2\boldsymbol{I})^{-1}$ and $\boldsymbol{P} = \mathbb{E}[\boldsymbol{xx}^\dagger]$. The computational complexity of the ZF and the MMSE detectors is mainly determined by matrix inversions and multiplications, which is of order $\mathcal{O}((NM)^3)$. Since we may assume $\text{tr}(\boldsymbol{H}^{(k,l)}\boldsymbol{H}^{(k,l)\dagger}) \ll \text{tr}(\boldsymbol{H}^{(k,k)}\boldsymbol{H}^{(k,k)\dagger})$ for $k \neq l$, the size of the matrices whose inverse are needed can be reduce to $M \times M$ by applying MMSE in a mode-wise manner. Indeed, the k-th entry $\boldsymbol{x}^{(k)}$ of $\boldsymbol{x}$ can be estimated as $\widehat{\boldsymbol{x}}^{(k)} = \boldsymbol{W}^{(k)}\boldsymbol{y}^{(k)}$, where

$$\boldsymbol{W}^{(k)} = \boldsymbol{P}^{(k)}\boldsymbol{H}^{(k,k)\dagger}\left(\boldsymbol{H}^{(k,k)}\boldsymbol{P}^{(k)}\boldsymbol{H}^{(k,k)\dagger} + \sigma^2\boldsymbol{I}\right)^{-1}, \quad (6)$$

and $\boldsymbol{P}^{(k)} = \mathbb{E}[\boldsymbol{x}^{(k)}\boldsymbol{x}^{(k)\dagger}]$. When Tx and Rx UCAs are ideally aligned (i.e., $\theta_x = \theta_y = \theta_z = 0$), the MMSE matrix $\boldsymbol{W}$ can be written as $\boldsymbol{W} = \text{diag}(\boldsymbol{W}^{(0)}, \boldsymbol{W}^{(1)}, \ldots, \boldsymbol{W}^{(N-1)})$, and the computational complexity of the mode-wise MMSE detector becomes $\mathcal{O}(NM^3)$. In this case, the OAM-MIMO channel can be thought of as a collection of N parallel $M\times M$ LOS-MIMO channels.

B. CHANNEL CAPACITY

We next investigate the transmission capacity of the OAM-MIMO channel presented in the previous subsection. When Tx and Rx multi-ring UCAs are ideally aligned, the Shannon capacity of the OAM-MIMO channel is given as follows:

$$\mathcal{C} = \sum_{k=0}^{N-1}\sum_{i=0}^{M-1}\log_2\left(1 + \lambda_i^{(k)}\frac{P_i^{(k)}}{\sigma^2}\right), \qquad (7)$$

where $(\lambda_0^{(k)}, \lambda_1^{(k)}, \ldots, \lambda_{M-1}^{(k)})$ denotes the eigenvalues of $\boldsymbol{H}^{(k,k)\dagger}\boldsymbol{H}^{(k,k)}$ and the sum of all the available power P is assumed to be distributed across the channels, according to the water-filling principle, such that $\sum_{k=0}^{N-1}\sum_{i=0}^{M-1}P_i^{(k)} = P$. Note that the capacity of (7) is achieved when the perfect channel state information (CSI) is known at the transmitter and the transmitted signal $\boldsymbol{x}^{(k)}$ is in the form of $\boldsymbol{x}^{(k)} = \boldsymbol{V}^{(k)}\boldsymbol{u}^{(k)}$, where $\boldsymbol{u}^{(k)}$ is uncorrelated Gaussian signal vector with $\mathbb{E}[\boldsymbol{u}^{(k)}\boldsymbol{u}^{(k)\dagger}] = \text{diag}(P_0^{(k)}, P_1^{(k)}, \ldots, P_{M-1}^{(k)})$ and $\boldsymbol{V}^{(k)}$ is the right singular matrix of $\boldsymbol{H}^{(k,k)}$. If the CSI is not known at the transmitter, a simpler transmit scheme with equal power allocation ($\boldsymbol{P}^{(k)} = P/(NM)\boldsymbol{I}$) yields

$$\overline{\mathcal{C}} = \sum_{k=0}^{N-1}\sum_{i=0}^{M-1}\log_2\left(1 + \lambda_i^{(k)}\frac{P}{NM\sigma^2}\right), \qquad (8)$$

which in general is inferior to (7) as power is wasted in eigenchannels with zero eigenvalue. It is known that if all eigenvalues are non-zero, (8) is asymptotically optimal at high signal-to-noise ratio (SNR).

We next consider the capacity for OAM-MIMO channel with the mode-wise MMSE estimation discussed in the previous subsection. The signal-to-interference-plus-noise ratio (SINR) for each MMSE estimated symbol $\widehat{x}_i^{(k)}$ of $\widehat{\boldsymbol{x}}^{(k)}$ is given by

$$\text{SINR}(\widehat{x}_i^{(k)}) = (\beta_i^{(k)})^2 \left(\alpha_i^{(k)} - (\beta_i^{(k)})^2 \right)^{-1}, \qquad (9)$$

where $\alpha_i^{(k)}$ and $\beta_i^{(k)}$ denote, respectively, diagonal entries of $\mathbb{E}[\widehat{\boldsymbol{x}}^{(k)}\widehat{\boldsymbol{x}}^{(k)\dagger}] (= \boldsymbol{W}^{(k)}\mathbb{E}[\boldsymbol{y}^{(k)}\boldsymbol{y}^{(k)\dagger}]\boldsymbol{W}^{(k)\dagger})$ and $\boldsymbol{W}^{(k)}\boldsymbol{H}^{(k,k)}$. The mode-wise MMSE capacity of the OAM-MIMO transmission model discussed in the previous subsection can be defined as follows:

$$\max_{\sum_{k=0}^{N-1} \text{tr}(\boldsymbol{P}^{(k)}) \leq P} \sum_{k=0}^{N-1} \sum_{i=0}^{M-1} \log_2 \left(1 + \frac{(\beta_i^{(k)})^2}{\alpha_i^{(k)} - (\beta_i^{(k)})^2} \right). \qquad (10)$$

Similarly to (8), the policy of equal power allocation yields

$$\text{SINR}(\widehat{x}_i^{(k)}) = \beta_i^{(k)}(1 - \beta_i^{(k)})^{-1} \qquad (11)$$

and

$$\mathcal{C}_{\text{MMSE}} = -\sum_{k=0}^{N-1} \sum_{i=0}^{M-1} \log_2 \left(1 - \beta_i^{(k)} \right). \qquad (12)$$

The mode-wise MMSE capacity of (12) approaches (8) when Tx and Rx UCAs are well-arranged, taking into account transmission distance and carrier frequency.

Example 1 (Single-ring UCA): As a typical example, let us consider an OAM-MIMO transmission system with single-ring (8, 1)-UCAs. The system consists of eight transmission modes, each of which can be thought of as a single-input and single-output (SISO) channel. We evaluate its channel capacity by (7) and (12) under the following conditions, which were selected with reference to those of the field trial experiment reported in [7]: (a) the Tx-Rx separation distance d is 40 m; (b) the carrier frequency f_c is 84.5 GHz (E-band); (c) the radius of the UCA is 0.265 m. Fig. 2 (Left) shows $\mathcal{C}$ and $\mathcal{C}_{\text{MMSE}}$ for each transmission mode[1] $-4 < k \leq 4$, where the SNR P/σ^2 was set to 35 dB and misalignments θ_x, θ_y and θ_z were assumed to be zero. The Shannon capacity $\mathcal{C}$ was computed by the water-filling method while $\mathcal{C}_{\text{MMSE}}$ was computed under the assumption of equal power allocation. In the setting of (a), (b) and (c), the approximation $\mathcal{C} \approx \mathcal{C}_{\text{MMSE}}$ holds for a wide range of SNR. The OAM modes of $k = \pm 2$ have the highest capacity and the other modes have almost the same capacity. Fig. 2 (Right) shows the mode-wise MMSE capacity versus z-axis misalignment θ_z, where θ_x and θ_y were fixed to zero. It should be noted that the MMSE capacities of all OAM-modes almost uniformly decrease with increasing misalignment angle.

Example 2 (Multi-ring UCA): Next we exhibit two examples of multi-ring UCA-based OAM-MIMO transmission

FIGURE 2. Channel Capacity for OAM-MIMO system with (8, 1)-UCAs. Left: Comparison of $\mathcal{C}$ and $\mathcal{C}_{\text{MMSE}}$ for each mode; Right: $\mathcal{C}_{\text{MMSE}}$ versus θ_z (rad).

FIGURE 3. Channel Capacity for OAM-MIMO system with (8, 2)-UCAs. Left: Comparison of $\mathcal{C}$ and $\mathcal{C}_{\text{MMSE}}$ for each mode; Right: $\mathcal{C}_{\text{MMSE}}$ versus θ_z (rad).

systems. The first one equips with (8, 2)-UCAs, and thus there are eight transmission modes, each of which can be thought of as a 2×2 LOS-MIMO channel. Fig. 3 (Left) shows its capacities $\mathcal{C}$ and $\mathcal{C}_{\text{MMSE}}$ at an SNR (P/σ^2) of 35 dB, where the power allocation method used was the same as in Example 1. Here, the transmission distance d and the carrier frequency f_c were set to 100 m and 130 GHz (D-band), and the outer and inner radiuses of the (8, 2)-UCA were set to 0.48 m and 0.30 m, respectively. These radiuses were optimized to maximize the Shannon capacity $\mathcal{C}$ subject to constraints on d, f_c and maximum antenna size. As can be seen from Fig. 3 (Left), the difference between $\mathcal{C}$ and $\mathcal{C}_{\text{MMSE}}$ is around 5.38 bps/Hz at SNR = 35 dB. The difference is almost constant for a wide range of SNR and is mainly due to the fact that the component subchannels in the 2×2 LOS-MIMO channels are not orthogonally multiplexed. More precisely, the columns of the channel matrix $\boldsymbol{H}^{(k,k)}$ are far from orthogonal for $k = 0, \pm 2, 4$. Fig. 3 (Right) shows the MMSE capacity for each mode k versus z-axis misalignment θ_z. Similarly to the case of Fig. 2 (Right), the MMSE capacities of all OAM-modes almost uniformly decrease with increasing misalignment angle. The next example is a case of (8, 3)-UCA. In this case, there are eight transmission mode, each of which can be thought of as a 3×3 LOS-MIMO channel. Similarly to Figs. 2 and 3, Fig. 4 shows $\mathcal{C}$ and $\mathcal{C}_{\text{MMSE}}$ at SNR = 35 dB, where d and f_c were set to 140 m and 157 GHz, and the radiuses of (8, 3)-UCA were set to 0.80 m, 0.49 m and 0.33 m, respectively. While the radiuses of the (8, 3)-UCA are optimized to maximize $\mathcal{C}$ under some constraints, the component subchannels in the 3×3 LOS-MIMO channels are not orthogonally multiplexed. With regard to the sensitivity of

---

[1] By convention, we use integers $-N/2 < k \leq N/2$ as OAM-mode indices rather than $0 \leq k < N$, which are congruent modulo N each other; and for simplicity, we denote by $\mathcal{C}^{(k)}$ the Shannon capacity of mode k and by $\mathcal{C}_{\text{MMSE},i}^{(k)}$ the MMSE capacity of the i-th channel in mode k.
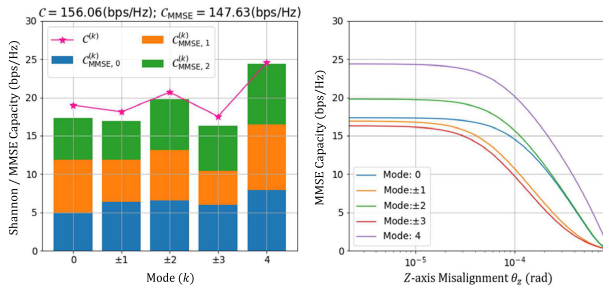
FIGURE 4. Channel Capacity for OAM-MIMO system with (8, 3)-UCAs. Left: Comparison of $\mathcal{C}$ and $\mathcal{C}_{MMSE}$ for each mode; Right: $\mathcal{C}_{MMSE}$ versus θ_z (rad).

antenna misalignment, the same tendency as noted above can be observed from Fig. 4.

The above Examples 1 and 2 show typical use of OAM-MIMO multiplexing for high capacity, PTP, LOS microwave/mm-wave radio systems [4]–[7], and these examples will be used in later sections as simulation settings to demonstrate performance of our learning-based OAM-MIMO detection algorithms.

III. LEARNING-BASED SIGNAL DETECTION
This section discusses and presents learning-based OAM-MIMO detection algorithms.

A. UNFOLDING ADMM FOR OAM-MIMO DETECTION
In the following discussions, we will use the same notation y, x and H to denote its real-valued matrix representation:

$$\begin{bmatrix} \Re(y) \\ \Im(y) \end{bmatrix}, \begin{bmatrix} \Re(x) \\ \Im(x) \end{bmatrix}, \begin{bmatrix} \Re(H) & -\Im(H) \\ \Im(H) & \Re(H) \end{bmatrix}.$$

Given channel matrix $H \in \mathbb{R}^{2NM \times 2NM}$ and received vector $y \in \mathbb{R}^{2NM}$, the original information vector x can be estimated by solving the following optimization problem:

$$\widehat{x} = \arg\min_{x \in \mathcal{S}^{2NM}} \|y - Hx\|_2^2, \tag{13}$$

where $\mathcal{S}$ denotes the set of I/Q values for signal constellation points, i.e., we may assume $\mathcal{S} = \{\pm 1, \pm 3, \ldots, \pm(2Q+1)\}$ for some appropriate non-negative integer Q. Unfortunately, the constraint of $x \in \mathcal{S}^{2NM}$ makes (13) computationally hard to solve, and the MMSE estimate discussed in Section II is a solution of the problem obtained by relaxing the constraint to $x \in \mathbb{R}^{2NM}$ and by adding an regularization term to the objective function. The MMSE estimate is known as a computationally efficient approximate solution of (13).

In the following, we consider the problem of (13) in the ADMM framework [22]. By introducing an auxiliary vector variable z, one can rewrite (13) as follows [21]:

$$\text{minimize } \frac{1}{2}\|y - Hx\|_2^2 + \mathbb{I}_{\mathcal{S}}(z),$$
$$\text{subject to } z = x, \tag{14}$$

where $\mathbb{I}_{\mathcal{S}}(z)$ is defined by

$$\mathbb{I}_{\mathcal{S}}(z) = \begin{cases} 0 & \text{if } z \in \mathcal{S}^{2NM}, \\ +\infty & \text{otherwise.} \end{cases} \tag{15}$$

The augmented Lagrangian function for (14) is written in scaled form as

$$\mathcal{L}_{\rho}(x, z, v) = \frac{1}{2}\|y - Hx\|_2^2 + \mathbb{I}_{\mathcal{S}}(z)$$
$$+ \frac{1}{2}\|\rho \circ (z - x + v)\|_2^2 - \frac{1}{2}\|\rho \circ v\|_2^2, \tag{16}$$

where v is the scaled dual variable vector and ρ is the penalty parameter vector. Then an ADMM-based algorithm for solving (14) can be described by the following iterations:

$$x_{\ell+1} = (H^T H + D_{\rho})^{-1}(H^T y + D_{\rho}(z_{\ell} + v_{\ell})), \tag{17}$$
$$z_{\ell+1} = \Pi_{\mathcal{S}}(x_{\ell+1} - v_{\ell}), \tag{18}$$
$$v_{\ell+1} = v_{\ell} + \eta \circ (z_{\ell+1} - x_{\ell+1}), \tag{19}$$

where ℓ denotes the iteration number; D_{ρ} denotes the diagonal matrix $\text{diag}(\rho \circ \rho)$; $\Pi_{\mathcal{S}}(\cdot)$ is the entry-wise projection onto the set $\mathcal{S}$; and η is an update rate vector. Note that if $\mathcal{S} = \{\pm 1, \pm 3, \ldots, \pm(2Q+1)\}$ for some Q, $\Pi_{\mathcal{S}}(x)$ can be given as

$$\Pi_{\mathcal{S}}(x) = \sum_{\iota=-Q}^{Q} \text{sign}(x - 2\iota), \tag{20}$$

where $\text{sign}(x) = 1$ if $x \geq 0$ and -1 otherwise.

Similarly to the ADMM-Net presented in [21], [23], we can construct a network with multiple layers by unfolding the iterations of (17)–(19) into layers and replacing parameters ρ and η with layer-wise parameters which can be trained using a gradient-descent method. More precisely, the unfolded ADMM network consists of L layers each with the same structure and the ℓ-th layer corresponds to the ℓ-th iteration of (17)–(19) for $\ell = 0, 1, \ldots, L-1$. Each layer inputs $(x_{\ell}, z_{\ell}, v_{\ell})$ and outputs $(x_{\ell+1}, z_{\ell+1}, v_{\ell+1})$, and the input-output relation can be expressed as follows:

$$x_{\ell+1} = (H^T H + D_{\rho_{\ell}})^{-1}(H^T y + D_{\rho_{\ell}}(z_{\ell} + v_{\ell})), \tag{21}$$
$$z_{\ell+1} = \psi_{\mathcal{S}, t_{\ell}}(x_{\ell+1} - v_{\ell}), \tag{22}$$
$$v_{\ell+1} = v_{\ell} + \eta_{\ell} \circ (z_{\ell+1} - x_{\ell+1}), \tag{23}$$

where ρ_{ℓ} and η_{ℓ} are layer-dependent learnable parameters and $\psi_{\mathcal{S}, t_{\ell}}(x)$ is an entry-wise non-linear function with learnable parameter t_{ℓ}. The non-linear function $\psi_{\mathcal{S}, t}(x)$ corresponding to (20) can be given as [19], [20]:

$$\psi_{\mathcal{S}, t}(x) = \sum_{\iota=-Q}^{Q} \left(-1 + \frac{R(x - 2\iota + t)}{t} - \frac{R(x - 2\iota - t)}{t} \right), \tag{24}$$

where $R(x)$ denotes the ReLU function $\max\{x, 0\}$. Note that the function $\psi_{\mathcal{S}, t}(x)$ approaches $\Pi_{\mathcal{S}}(x)$ as t goes to zero. The computational complexity of the above ADMM-based detector is governed by matrix inversions and multiplications, which is of order $\mathcal{O}((NM)^3 L)$. Recall here that it may be assumed $\text{tr}(H^{(k,l)} H^{(k,l)T}) \ll \text{tr}(H^{(k,k)} H^{(k,k)T})$ for $k \neq l$. Similarly to the MMSE case mentioned in Section II-B, the size of the matrices whose inverse are needed can be

reduce to $M \times M$ by applying ADMM in a mode-wise manner. In addition, by performing parallel interference cancellation (PIC) layer by layer, the mode-wise ADMM will enhance the tolerance to inter-mode interference due to antenna misalignment. The resulting ADMM-PIC detector consists of multiple layers, and at the first layer, the k-th entry $(\boldsymbol{x}_1^{(k)}, \boldsymbol{z}_1^{(k)}, \boldsymbol{v}_1^{(k)})$ of $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v})$ is estimated from $\boldsymbol{H}^{(k,k)}$ and $\boldsymbol{y}^{(k)}$ for all k. Then, at a subsequent layer $\ell + 1$, they are refined by using $\boldsymbol{y}^{(k)} - \boldsymbol{\gamma}_\ell^{(k)} \circ \sum_{l=0,l\neq k}^{N-1} \boldsymbol{H}^{(k,l)}\boldsymbol{x}_\ell^{(l)}$ rather than $\boldsymbol{y}^{(k)}$, where $\boldsymbol{\gamma}_\ell^{(k)}$ is a learnable parameter vector.

The processing of the ADMM-PIC detector can be summarized as the following pseudocode, which we will refer to as Algorithm 1.

Algorithm 1 Unfolded ADMM-PIC

Input: $(\boldsymbol{y}^{(0)}, \boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(N-1)})$ and $(\boldsymbol{H}^{(k,l)})_{0 \leq k,l < N}$
Output: $(\widehat{\boldsymbol{x}}^{(0)}, \widehat{\boldsymbol{x}}^{(1)}, \ldots, \widehat{\boldsymbol{x}}^{(N-1)})$
1: $(\boldsymbol{x}_0^{(k)}, \boldsymbol{z}_0^{(k)}, \boldsymbol{v}_0^{(k)}) \leftarrow (\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0})$ for $0 \leq k < N$
2: **for** $\ell = 0, 1, \ldots, L-1$ **do**
3:     **for** $k = 0, 1, \ldots, N-1$ **do**
4:         $\widetilde{\boldsymbol{y}}_\ell^{(k)} \leftarrow \boldsymbol{y}^{(k)} - \boldsymbol{\gamma}_\ell^{(k)} \circ \sum_{l=0,l\neq k}^{N-1} \boldsymbol{H}^{(k,l)}\boldsymbol{x}_\ell^{(l)}$
5:     **end for**
6:     **for** $k = 0, 1, \ldots, N-1$ **do**
7:         $\boldsymbol{x}_{\ell+1}^{(k)} \leftarrow (\boldsymbol{H}^{(k,k)T}\boldsymbol{H}^{(k,k)} + \boldsymbol{D}_{\boldsymbol{\rho}_\ell^{(k)}})^{-1}$
8:         $\cdot (\boldsymbol{H}^{(k,k)T}\widetilde{\boldsymbol{y}}_\ell^{(k)} + \boldsymbol{D}_{\boldsymbol{\rho}_\ell^{(k)}}(\boldsymbol{z}_\ell^{(k)} + \boldsymbol{v}_\ell^{(k)}))$
9:         $\boldsymbol{z}_{\ell+1}^{(k)} \leftarrow \psi_{\mathcal{S},\boldsymbol{t}_\ell^{(k)}}(\boldsymbol{x}_{\ell+1}^{(k)} - \boldsymbol{v}_\ell^{(k)})$
10:       $\boldsymbol{v}_{\ell+1}^{(k)} \leftarrow \boldsymbol{v}_\ell^{(k)} + \boldsymbol{\eta}_\ell^{(k)} \circ (\boldsymbol{z}_{\ell+1}^{(k)} - \boldsymbol{x}_{\ell+1}^{(k)})$
11:     **end for**
12: **end for**
13: $\widehat{\boldsymbol{x}}^{(k)} \leftarrow \psi_{\mathcal{S},\boldsymbol{t}_L^{(k)}}(\boldsymbol{x}_L^{(k)})$ for $0 \leq k < N$

The computational complexity of Algorithm 1 is of order $\mathcal{O}(NM^3L + (NM)^2L)$. For each layer $\ell = 0, 1, \ldots, L-1$, the learnable parameters in Algorithm 1 are $\boldsymbol{\rho}_\ell^{(k)}, \boldsymbol{t}_\ell^{(k)}, \boldsymbol{\eta}_\ell^{(k)}$ and $\boldsymbol{\gamma}_\ell^{(k)}$ for $k = 0, 1, \ldots, N-1$, and thus the number of parameters to be learned in each layer is $8NM$. This is much less than the number used in the DetNet [19], [20] which is determined not only by the number of antenna elements but also by hyperparameters such as the number of hidden neurons. For details on the learnable parameters in the DetNet, refer to Appendix A-B. The loss function we use for training these parameters is given by

$$E(\boldsymbol{x}, \widehat{\boldsymbol{x}}) = \frac{1}{2}\sum_{k=0}^{N-1} \|\boldsymbol{x}^{(k)} - \widehat{\boldsymbol{x}}^{(k)}\|_2^2, \quad (25)$$

where $\boldsymbol{x} = (\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N-1)})$ is the original information and $\widehat{\boldsymbol{x}} = (\widehat{\boldsymbol{x}}^{(0)}, \widehat{\boldsymbol{x}}^{(1)}, \ldots, \widehat{\boldsymbol{x}}^{(N-1)})$ is its estimate obtained by Algorithm 1.

The number of matrix inversion operations of size $M \times M$ required in Algorithm 1 is NL, which can be reduced to N by replacing L parameters $\boldsymbol{\rho}_0^{(k)}, \boldsymbol{\rho}_1^{(k)}, \ldots, \boldsymbol{\rho}_{L-1}^{(k)}$ with a single common parameter $\boldsymbol{\rho}^{(k)}$. Similarly, while the estimation

accuracy may be degenerated, the computational complexity required for training and testing the ADMM-PIC network can be further reduced by replacing parameters $\boldsymbol{t}_\ell^{(k)}, \boldsymbol{\eta}_\ell^{(k)}, \boldsymbol{\gamma}_\ell^{(k)}$ with appropriate predetermined values. For example, $\boldsymbol{\eta}_\ell^{(k)}$ and $\boldsymbol{\gamma}_\ell^{(k)}$ may be replaced with all-one vector, and $\boldsymbol{t}_\ell^{(k)}$ may be replaced with all-zero vector (which means replacing the non-linear function $\psi_{\mathcal{S},\boldsymbol{t}_\ell^{(k)}}$ with $\Pi_{\mathcal{S}}$).

The following algorithm (referred to as Algorithm 2) is a simplified version of Algorithm 1, in which the total number of parameters to be learned is $2NM(1+L)$ and the number of the matrix inversion operations is N.

Algorithm 2 Simplified Version of Algorithm 1

Input: $(\boldsymbol{y}^{(0)}, \boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(N-1)})$ and $(\boldsymbol{H}^{(k,l)})_{0 \leq k,l < N}$
Output: $(\widehat{\boldsymbol{x}}^{(0)}, \widehat{\boldsymbol{x}}^{(1)}, \ldots, \widehat{\boldsymbol{x}}^{(N-1)})$
1: $(\boldsymbol{x}_0^{(k)}, \boldsymbol{z}_0^{(k)}, \boldsymbol{v}_0^{(k)}) \leftarrow (\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0})$ for $0 \leq k < N$
2: $\boldsymbol{G}^{(k)} \leftarrow (\boldsymbol{H}^{(k,k)T}\boldsymbol{H}^{(k,k)} + \boldsymbol{D}_{\boldsymbol{\rho}^{(k)}})^{-1}$ for $0 \leq k < N$
3: **for** $\ell = 0, 1, \ldots, L-1$ **do**
4:     **for** $k = 0, 1, \ldots, N-1$ **do**
5:         $\widetilde{\boldsymbol{y}}_\ell^{(k)} \leftarrow \boldsymbol{y}^{(k)} - \boldsymbol{\gamma}_\ell^{(k)} \circ \sum_{l=0,l\neq k}^{N-1} \boldsymbol{H}^{(k,l)}\boldsymbol{x}_\ell^{(l)}$
6:     **end for**
7:     **for** $k = 0, 1, \ldots, N-1$ **do**
8:         $\boldsymbol{x}_{\ell+1}^{(k)} \leftarrow \boldsymbol{G}^{(k)}(\boldsymbol{H}^{(k,k)T}\widetilde{\boldsymbol{y}}_\ell^{(k)} + \boldsymbol{D}_{\boldsymbol{\rho}^{(k)}}(\boldsymbol{z}_\ell^{(k)} + \boldsymbol{v}_\ell^{(k)}))$
9:         $\boldsymbol{z}_{\ell+1}^{(k)} \leftarrow \Pi_{\mathcal{S}}(\boldsymbol{x}_{\ell+1}^{(k)} - \boldsymbol{v}_\ell^{(k)})$
10:       $\boldsymbol{v}_{\ell+1}^{(k)} \leftarrow \boldsymbol{v}_\ell^{(k)} + (\boldsymbol{z}_{\ell+1}^{(k)} - \boldsymbol{x}_{\ell+1}^{(k)})$
11:     **end for**
12: **end for**
13: $\widehat{\boldsymbol{x}}^{(k)} \leftarrow \Pi_{\mathcal{S}}(\boldsymbol{x}_L^{(k)})$ for $0 \leq k < N$

Note here that lines 7–11 of the pseudocode of Algorithm 2 corresponds to the ADMM-Net presented in [21]. More precisely, the ADMM-Net [21] can be obtained from lines 7–11 of the above pseudocode by replacing $\Pi_{\mathcal{S}}(\cdot)$ in line 9 with $\psi_{\mathcal{S},\boldsymbol{t}}(\cdot)$ with learnable parameter $\boldsymbol{t}$. We also note that if, for $k = 0, 1, \ldots, N-1$, $\boldsymbol{\rho}^{(k)}$ is set as $\boldsymbol{D}_{\boldsymbol{\rho}^{(k)}} = (NM\sigma^2/P)\boldsymbol{I}$, then the mode-wise MMSE estimate of $\boldsymbol{x}^{(k)}$ is obtained in the first iteration of the above algorithm, i.e., $\boldsymbol{x}_\ell^{(k)} = \boldsymbol{W}^{(k)}\boldsymbol{y}^{(k)}$ when $\ell = 1$. It is thus expected that the subsequent iterations of Algorithm 2 yield more accurate estimate than the MMSE method.

B. NETWORK TRAINING

The set of learnable parameters in the ADMM-PIC network of Algorithm 1 is $\{(\boldsymbol{\rho}_\ell^{(k)}, \boldsymbol{t}_\ell^{(k)}, \boldsymbol{\eta}_\ell^{(k)}, \boldsymbol{\gamma}_\ell^{(k)}) \mid 0 \leq k < N, 0 \leq \ell < L\}$. These parameters can be learned by using stochastic gradient descent and back-propagation. We mention here the gradient computation for back propagation of the ADMM-PIC network. The following notation will be used in this section: For a vector $\boldsymbol{u} = (u_i)$, $\partial E/\partial \boldsymbol{u}$ denotes the gradient (column) vector of the loss function of (25) w. r. t. $\boldsymbol{u}$, and for two vectors $\boldsymbol{v} = (v_i)$ and $\boldsymbol{w} = (w_i)$, $\partial \boldsymbol{v}/\partial \boldsymbol{w}$ denotes a matrix whose (i, j)-th entry is $\partial v_j/\partial w_i$.

Then, for $0 \le k < N$ and $0 \le \ell < L$, the gradient vector $\partial E / \partial \boldsymbol{\rho}_\ell^{(k)}$ is computed as follows:

$$
\begin{aligned}
\frac{\partial E}{\partial \boldsymbol{\rho}_\ell^{(k)}} &= \frac{\partial \boldsymbol{x}_{\ell+1}^{(k)}}{\partial \boldsymbol{\rho}_\ell^{(k)}} \cdot \frac{\partial E}{\partial \boldsymbol{x}_{\ell+1}^{(k)}} \\
&= \operatorname{diag}\big(2\boldsymbol{\rho}_\ell^{(k)} \circ (\boldsymbol{z}_\ell^{(k)} + \boldsymbol{v}_\ell^{(k)} - \boldsymbol{x}_{\ell+1}^{(k)})\big) \cdot \boldsymbol{q}_\ell^{(k)},
\end{aligned}
\tag{26}
$$

where $\boldsymbol{q}_\ell^{(k)}$ is defined as

$$
\boldsymbol{q}_\ell^{(k)} = \big(\boldsymbol{H}^{(k,k)T} \boldsymbol{H}^{(k,k)} + \boldsymbol{D}_{\boldsymbol{\rho}_\ell^{(k)}}\big)^{-1} \frac{\partial E}{\partial \boldsymbol{x}_{\ell+1}^{(k)}}.
\tag{27}
$$

For the parameter vector $\boldsymbol{t}_{\ell-1}^{(k)}$, we have

$$
\begin{aligned}
\frac{\partial E}{\partial \boldsymbol{t}_{\ell-1}^{(k)}} &= \frac{\partial \boldsymbol{z}_\ell^{(k)}}{\partial \boldsymbol{t}_{\ell-1}^{(k)}} \cdot \frac{\partial \boldsymbol{x}_{\ell+1}^{(k)}}{\partial \boldsymbol{z}_\ell^{(k)}} \cdot \frac{\partial E}{\partial \boldsymbol{x}_{\ell+1}^{(k)}} \\
&= \operatorname{diag}\big((1 + \boldsymbol{\eta}_{\ell-1}^{(k)}) \circ \varphi_{\mathcal{S},\boldsymbol{t}_{\ell-1}^{(k)}}(\boldsymbol{x}_\ell^{(k)} - \boldsymbol{v}_{\ell-1}^{(k)})\big) \\
&\quad \cdot \boldsymbol{D}_{\boldsymbol{\rho}_\ell^{(k)}} \cdot \boldsymbol{q}_\ell^{(k)},
\end{aligned}
\tag{28}
$$

where $\varphi_{\mathcal{S},\boldsymbol{t}}(\boldsymbol{x})$ denotes $(\partial \psi_{\mathcal{S},t_i}(x_i) / \partial t_i)$. Similarly, for the parameter vectors $\boldsymbol{\eta}_{\ell-1}^{(k)}$ and $\boldsymbol{\gamma}_\ell^{(k)}$, we have

$$
\begin{aligned}
\frac{\partial E}{\partial \boldsymbol{\eta}_{\ell-1}^{(k)}} &= \frac{\partial \boldsymbol{v}_\ell^{(k)}}{\partial \boldsymbol{\eta}_{\ell-1}^{(k)}} \cdot \frac{\partial \boldsymbol{x}_{\ell+1}^{(k)}}{\partial \boldsymbol{v}_\ell^{(k)}} \cdot \frac{\partial E}{\partial \boldsymbol{x}_{\ell+1}^{(k)}} \\
&= \operatorname{diag}\big(\boldsymbol{z}_\ell^{(k)} - \boldsymbol{x}_\ell^{(k)}\big) \cdot \boldsymbol{D}_{\boldsymbol{\rho}_\ell^{(k)}} \cdot \boldsymbol{q}_\ell^{(k)},
\end{aligned}
\tag{29}
$$

$$
\begin{aligned}
\frac{\partial E}{\partial \boldsymbol{\gamma}_\ell^{(k)}} &= \frac{\partial \boldsymbol{x}_{\ell+1}^{(k)}}{\partial \boldsymbol{\gamma}_\ell^{(k)}} \cdot \frac{\partial E}{\partial \boldsymbol{x}_{\ell+1}^{(k)}} \\
&= -\operatorname{diag}\Big(\sum_{l=0,l\ne k}^{N-1} \boldsymbol{H}^{(k,l)} \boldsymbol{x}_\ell^{(l)}\Big) \cdot \boldsymbol{H}^{(k,k)} \cdot \boldsymbol{q}_\ell^{(k)}.
\end{aligned}
\tag{30}
$$

Next, $\partial E / \partial \boldsymbol{x}_\ell^{(k)}$ can be computed as follows:

$$
\begin{aligned}
\frac{\partial E}{\partial \boldsymbol{x}_\ell^{(k)}} &= \sum_{l=0}^{N-1} \frac{\partial \boldsymbol{x}_{\ell+1}^{(l)}}{\partial \boldsymbol{x}_\ell^{(k)}} \cdot \frac{\partial E}{\partial \boldsymbol{x}_{\ell+1}^{(l)}} \\
&= \operatorname{diag}\Big((1 + \boldsymbol{\eta}_{\ell-1}^{(k)}) \circ \phi_{\mathcal{S},\boldsymbol{t}_{\ell-1}^{(k)}}(\boldsymbol{x}_\ell^{(k)} - \boldsymbol{v}_{\ell-1}^{(k)}) - \boldsymbol{\eta}_{\ell-1}^{(k)}\Big) \\
&\quad \cdot \boldsymbol{D}_{\boldsymbol{\rho}_\ell^{(k)}} \cdot \boldsymbol{q}_\ell^{(k)} - \sum_{l=0,l\ne k}^{N-1} \boldsymbol{H}^{(l,k)T} \boldsymbol{H}^{(l,l)} \boldsymbol{q}_\ell^{(l)},
\end{aligned}
\tag{31}
$$

where $\phi_{\mathcal{S},\boldsymbol{t}}(\boldsymbol{x})$ denotes $(\partial \psi_{\mathcal{S},t_i}(x_i) / \partial x_i)$. By (31), the partial derivatives $\partial E / \partial \boldsymbol{x}_\ell^{(k)}$ for $0 \le k < N$ are computed in the order of $\ell = L-1, L-2, \ldots, 1$. The gradients $\partial E / \partial \boldsymbol{\rho}_\ell^{(k)}$, $\partial E / \partial \boldsymbol{t}_{\ell-1}^{(k)}$, $\partial E / \partial \boldsymbol{\eta}_{\ell-1}^{(k)}$ and $\partial E / \partial \boldsymbol{\gamma}_\ell^{(k)}$ are then determined by (26), (28), (29), and (30), respectively.

Finally, we mention the training of the parameters $\boldsymbol{\gamma}_\ell^{(k)}$ and $\boldsymbol{\rho}^{(k)}$ in Algorithm 2. Note that the gradient w.r.t. the parameter $\boldsymbol{\rho}^{(k)}$ is computed as follows:

$$
\begin{aligned}
\frac{\partial E}{\partial \boldsymbol{\rho}^{(k)}} &= \sum_{\ell=0}^{L-1} \frac{\partial \boldsymbol{x}_{\ell+1}^{(k)}}{\partial \boldsymbol{\rho}^{(k)}} \cdot \frac{\partial E}{\partial \boldsymbol{x}_{\ell+1}^{(k)}} \\
&= \sum_{\ell=0}^{L-1} \operatorname{diag}\big(2\boldsymbol{\rho}^{(k)} \circ (\boldsymbol{z}_\ell^{(k)} + \boldsymbol{v}_\ell^{(k)} - \boldsymbol{x}_{\ell+1}^{(k)})\big) \cdot \overline{\boldsymbol{q}}_\ell^{(k)},
\end{aligned}
$$

where $\overline{\boldsymbol{q}}_\ell^{(k)} = \boldsymbol{G}^{(k)} \cdot \partial E / \partial \boldsymbol{x}_{\ell+1}^{(k)}$. We also note that $\partial E / \partial \boldsymbol{\gamma}_\ell^{(k)}$ and $\partial E / \partial \boldsymbol{x}_\ell^{(k)}$ can be computed in a similar manner as noted above.

IV. SIMULATION RESULTS

This section presents simulation results for the ADMM-based OAM-MIMO detectors presented in the previous section. We consider the UCA-based OAM-MIMO systems discussed in Examples 1 and 2 of Section II-B. In all our simulations, the equal transmission power allocation is assumed, and thus the SNR per transmit symbol E_s/N_0 is given by $P/(NM\sigma^2)$. We also assume that all information symbols are modulated using the Gray-coded 256 QAM with the constellation points $\mathcal{S} = \{\pm 1, \pm 3, \ldots, \pm 15\}$. For convenience, we define the following signal-to-interference ratio (SIR) for each mode k:

$$
\operatorname{SIR}(k) = \frac{\operatorname{tr}(\boldsymbol{H}^{(k,k)} \boldsymbol{H}^{(k,k)T})}{\sum_{l=0,l\ne k}^{N-1} \operatorname{tr}(\boldsymbol{H}^{(k,l)} \boldsymbol{H}^{(k,l)T})}.
\tag{32}
$$

We use the average of $\operatorname{SIR}(k)$ over $k = 0, 1, \ldots, N-1$ to measure the degree of inter-mode interference due to antenna misalignment.

Fig. 5 (Left) shows the flowchart of the training process of our ADMM-PIC network for OAM-MIMO detection. At each epoch of the training process, transmit signal sequence $\boldsymbol{x}$ of length equal to the batch size is uniformly randomly generated, and Gaussian noise signal sequence $\boldsymbol{n}$ of the same length is generated, where the SNR of each noise signal is chosen independently and uniformly from the predetermined range. In addition, the channel matrix $\boldsymbol{H}$ is generated under the conditions described in Examples 1 and 2, where the misalignment angles are chosen independently and randomly at each epoch so that the resulting SIR is distributed in the predetermined ranges depending on the type of UCA. Then, given the received signal sequence $\boldsymbol{y}$ and the channel matrix $\boldsymbol{H}$, the ADMM-PIC detector computes by Algorithm 1 (or 2) the estimate $\widehat{\boldsymbol{x}}$ of $\boldsymbol{x}$. The training procedure then computes the squared error between $\widehat{\boldsymbol{x}}$ and $\boldsymbol{x}$ and updates the parameters in the ADMM-PIC network by using back propagation as noted in Section III-B. In our simulations, the batch size and the number of epochs used for training were set to 2,000 and 10,000, respectively. The number of the layers of the ADMM-PIC network was set to $L = 5$, and the Adam optimizer [24] with a learning rate of 10^{-3} was used for training. Table 1 (upper) summarizes the training conditions for the ADMM-PIC networks.

Fig. 5 (Right) shows the flowchart of the testing process. Similarly to the training process, transmit signal sequence $\boldsymbol{x}$ and Gaussian noise signal sequence $\boldsymbol{n}$ are generated, where the SNR is fixed to the predetermined value. The channel matrix $\boldsymbol{H}$ is also generated in a similar manner to the training process, where the misalignment angles are chosen independently and randomly at each epoch so that the resulting SIR is approximately equal to the predetermined value. The testing procedure computes the estimate $\widehat{\boldsymbol{x}}$ of $\boldsymbol{x}$ from $\boldsymbol{y}$ and $\boldsymbol{H}$ by using Algorithm 1 (or 2) with the parameters determined by the
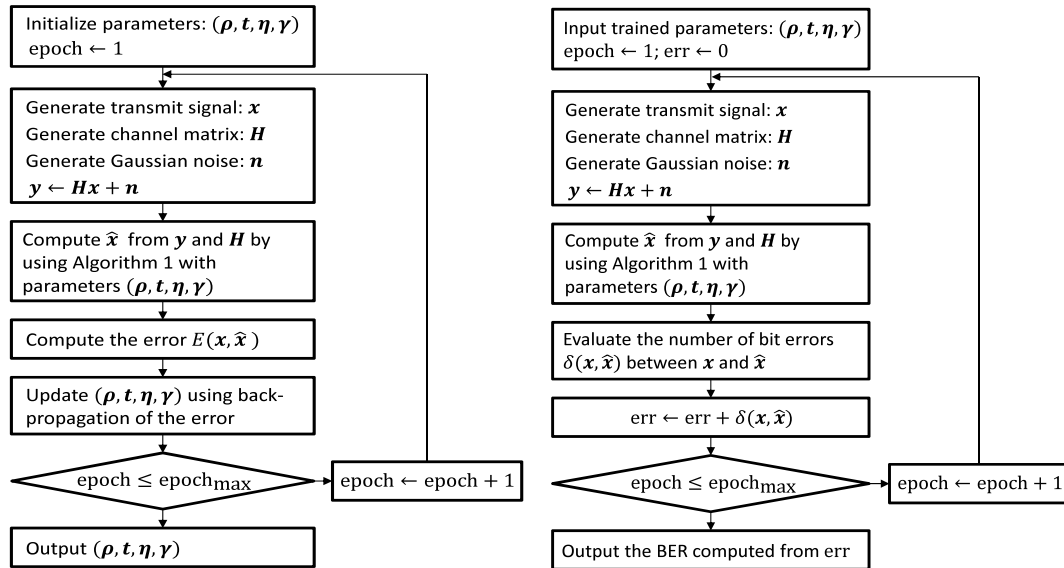
FIGURE 5. Flowcharts of the training process (Left) and the testing process (Right).

TABLE 1. Training conditions for ADMM-PIC (upper) and DetNet-PIC (lower).

		ADMM-PIC					
(N, M)	L	Number of parameters (Alg.1 / Alg.2)	Batch size	Number of epochs	Modulation scheme	SNR range (dB)	SIR range (dB)
$(8,1)$	5	320 / 96	2,000	10,000	256 QAM	$33 \sim 37$	$6 \sim 16$
$(8,2)$	5	640 / 192	2,000	10,000	256 QAM	$34 \sim 38$	$8 \sim 18$
$(8,3)$	5	960 / 288	2,000	10,000	256 QAM	$34 \sim 38$	$10 \sim 20$
		DetNet-PIC					
(N, M)	L	Number of parameters	Batch size	Number of epochs	Modulation scheme	SNR range (dB)	SIR range (dB)
$(8,1)$	5	3,408	2,000	50,000	256 QAM	$33 \sim 37$	$6 \sim 16$
$(8,2)$	5	12,944	2,000	50,000	256 QAM	$34 \sim 38$	$8 \sim 18$
$(8,3)$	5	28,624	2,000	50,000	256 QAM	$34 \sim 38$	$10 \sim 20$

above training process. These training and testing processes were implemented by TensorFlow [25] on NVIDIA GeForce RTX 2080 with 8 CPU cores and 2 GPUs each with 8 GB RAM.

The following figures (Figs. 6, 7, and 8) show the bit error rate (BER) performance of Algorithms 1 and 2 for the 256 QAM OAM-MIMO systems described in Examples 1 and 2. For purposes of comparison, these figures also show the BER performance of the MMSE-PIC and the DetNet-PIC detectors which are explained in the appendix. Similarly to the ADMM-PIC detector presented in Section III-A, the MMSE-PIC detector (resp. the DetNet-PIC detector) consists of iterations of the mode-wise MMSE (resp. DetNet [20]) process, in each of which the PIC process is integrated. The MMSE-PIC has no trainable parameters and thus the training process is not necessary. On the other hand, the DetNet-PIC has a large number of trainable parameters. In our simulations, while the training process for the DetNet-PIC was the same as that for the ADMM-PIC, the number of epochs for training the DetNet-PIC was five time as much as that for the ADMM-PIC. The training conditions for the DetNet-PIC are summarized in Table 1 (lower).

Fig. 6 (Left) plots the BER versus SNR per transmit symbol E_s/N_0 for the $(8, 1)$-UCA based OAM-MIMO system

described in Example 1. The BERs were computed with the above testing process, where the SIR was fixed to 6 dB and 10 dB while E_s/N_0 was varied from 6 dB to 30 dB. From Fig. 6 (Left), it can be seen that when SIR is relatively low (6 dB) the MMSE-PIC performs poorly in comparison to the others and the ADMM-PICs are slightly better than the DetNet-PIC in BER performance, while when SIR is relatively high (10 dB), all the OAM-MIMO detectors yield similar BER performance (the DetNet-PIC is slightly better than the others). To further observe the impact of inter-mode interference, Fig. 6 (Right) plots the BERs for various values of SIR, when E_s/N_0 was fixed to 23 dB and 26 dB. It can be seen that the ADMM-PIC and the DetNet-PIC detectors achieve better BER performance than the MMSE-PIC over a wide range of SIR, especially when $E_s/N_0 = 26$ dB. Note that when $E_s/N_0 = 23$ dB, the BER curves for the ADMM-PIC and the DetNet-PIC detectors exhibit error-floors at a BER of around 10^{-4} and that the performance of these detectors become almost the same as that of the MMSE-PIC in a relatively high SIR region (SIR $\geq$ 10 dB). As can be seen, the location of the error-floor depends on the setting of E_s/N_0, and it goes down as increasing E_s/N_0. From Fig. 6, it can be observed that in this case, where $(N, M) = (8, 1)$, both Algorithms 1 and 2 perform much better than the MMSE-PIC and
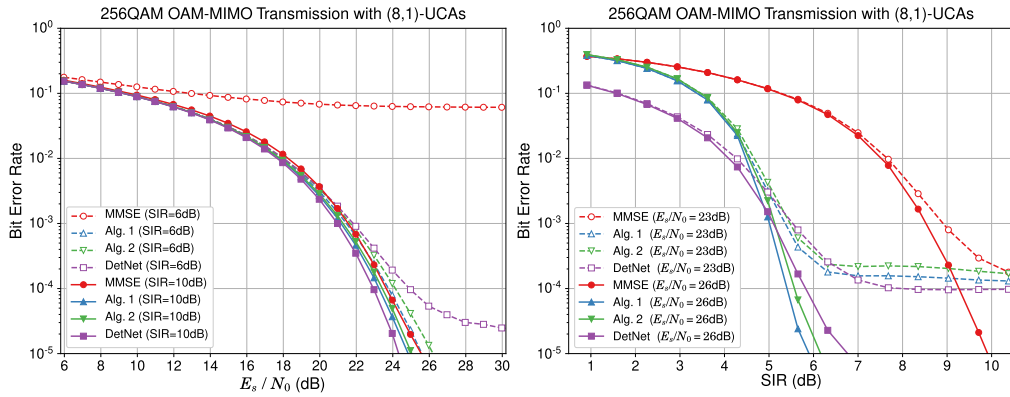
FIGURE 6. Bit error rate (BER) performance of MIMO detectors for 256 QAM OAM-MIMO transmission system with (8,1)-UCAs. Left: BER versus E_s/N_0, where SIR = 6 dB and 10 dB. Right: BER versus SIR, where E_s/N_0 = 23 dB and 26 dB.
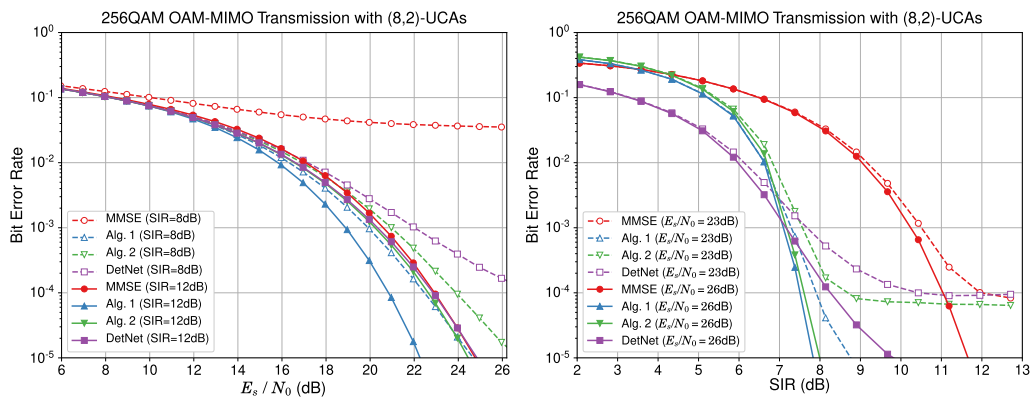

FIGURE 7. Bit error rate (BER) performance of MIMO detectors for 256 QAM OAM-MIMO transmission system with (8,2)-UCAs. Left: BER versus E_s/N_0, where SIR = 8 dB and 12 dB. Right: BER versus SIR, where E_s/N_0 = 23 dB and 26 dB.

that they achieve almost the same level of BER performance as the DetNet-PIC while requiring much smaller number of learnable parameters. From Table 1 the number of learnable parameters of Algorithms 1 and 2 are, respectively, only 9 percent and 3 percent of that of the DetNet-PIC.

Next we show the simulation results for OAM-MIMO systems with multi-ring UCAs. Fig. 7 shows the BER performance for the (8, 2)-UCA based OAM-MIMO system in Example 2. Fig. 7 (Left) plots the BER versus SNR per transmit symbol E_s/N_0, where the SIR was fixed to 8 dB and 12 dB while E_s/N_0 was varied from 6 dB to 26 dB. Similarly to the case of Fig. 6, while the ADMM-PIC detectors perform very well, the MMSE-PIC performs poorly when the power of inter-mode interference is relatively high (SIR = 8 dB), and the DetNet-PIC performs worse than the ADMM-PIC detectors. It can also be observed that the complete ADMM-PIC (Algorithm 1) performs much better than its simplified version (Algorithm 2), especially in relatively low interference power regime. Indeed, when SIR = 12 dB, Algorithm 1 provides more than 2 dB gain at a BER of 10^{-5} over Algorithm 2. This performance gain comes from the selection of learnable parameters. As noted in Section III-A, Algorithm 2 is inherently similar to the MMSE in a low interference power region and that the MMSE is far from

optimal when component sub-channels in an OAM-mode (viewed as a 2×2 LOS-MIMO channel) are not orthogonally multiplexed. Fig. 7 (Right) plots the BERs for various values of SIR, when E_s/N_0 was fixed to 23 dB and 26 dB. As can be seen, Algorithm 1 can achieve better BER performance than Algorithm 2 and the MMSE-PIC over a wide range of SIR. The DetNet-PIC performs better than the others at low SIR, but it performs worse than the ADMM-PICs at higher SIR. It can be observed that when $E_s/N_0 = 23$ dB, the improvements in BERs of Algorithm 2 and the DetNet-PIC are saturated at around SIR = 8 dB, and they become almost the same as the MMSE-PIC in a high SIR region (SIR $\geq$ 12 dB). Similarly to the case of Fig. 6 (Right), the location of the error-floor depends on the setting of E_s/N_0, and it goes down as increasing E_s/N_0.

Finally, Fig. 8 shows the simulation results for the OAM-MIMO system with (8, 3)-UCAs described in Example 2. Similarly to Figs. 6 and 7, this figure plots the BERs of the four detectors, where the SIR was fixed to 10 dB and 13 dB in the left figure and E_s/N_0 was fixed to 22 dB and 25 dB in the right figure. As can be seen from the figure, the complete ADMM-PIC (Algorithm 1) performs very well and outperforms the other three detectors for wide ranges of SIR and of E_s/N_0. The simplified ADMM-PIC
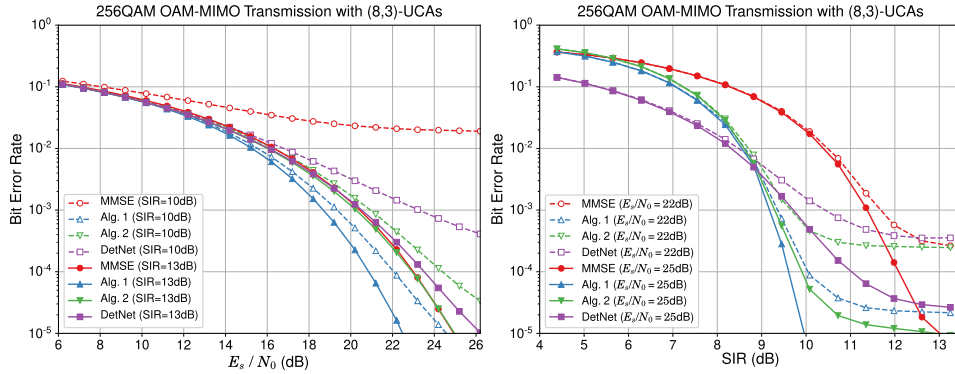
FIGURE 8. Bit error rate (BER) performance of MIMO detectors for 256 QAM OAM-MIMO transmission system with (8,3)-UCAs. Left: BER versus E_S/N_0, where SIR = 10 dB and 13 dB. Right: BER versus SIR, where E_S/N_0 = 22 dB and 25 dB.

(Algorithm 2) and the DetNet-PIC perform better than the MMSE-PIC at a low SIR, but they suffer from an error-floor at higher SIR region where they perform almost the same as the MMSE-PIC. The range of SIR where the DetNet-PIC performs better than the others is limited to a very low SIR region.

V. CONCLUSION

In this paper, we investigated the channel capacity for UCA-based OAM-MIMO systems and demonstrated using examples that the mode-wise MMSE capacity and the MMSE detector performance deteriorate seriously due to inter-mode interference that occurs when Tx and Rx UCAs are not ideally aligned. To address this issue, we then presented a learning, neural-like network-based signal detection method, which is derived by integrating unfolded ADMM with PIC. The number of parameters needed to be trained is only a linear order of the number of the antenna elements. We also presented a simplified version of the ADMM-PIC, where the number of parameters is reduced to around 30 percent of the complete version.

Our simulation results show that the proposed ADMM-PIC detector outperforms in BER performance the MMSE-based detector not only in the case of severe inter-mode interference but also in the case of low interference. The simulations also show that the proposed detector is comparable to or outperforms the DetNet-based detector with a smaller number of learnable parameters. The number of learnable parameters in the proposed detector is about 3–10 percent of that in the DetNet-based detector. While we focused in this paper on OAM-MIMO systems with UCAs, our work can be extended to other MIMO systems such as an LOS-MIMO system with uniform linear array antennas [15].

The next step for future work is to investigate and evaluate practical performance of the proposed method using an authentic set of data from real systems in actual physical environments. The effectiveness of our approach shown in this work was only evaluated through simulations of mathematical model, and at this point, this work is still in a simulation stage as with most prior work on this topic. The

effectiveness in a real channel environment remains to be demonstrated in future work.

APPENDIX A
MMSE-PIC AND DetNet-PIC

In this appendix, we explain the MMSE-PIC and the DetNet-PIC which were introduced and used in Section IV for the purpose of comparison.

A. MMSE-PIC

The MMSE-PIC detector consists of L stages, where L was set to 5 in the simulations of Section IV. At the first stage, the transmitted signal $x^{(k)}$ is estimated from the received signal $y^{(k)}$ and the mode-wise MMSE filter coefficient $W^{(k)}$ by $\widehat{x}^{(k)} = W^{(k)} y^{(k)}$, where $W^{(k)}$ is given as in (6). Then at a subsequent stage ℓ, each estimate is simply refined as

$$\widehat{x}_\ell^{(k)} = W^{(k)}\left(y^{(k)} - \sum_{l=0,l\neq k}^{N-1} H^{(k,l)}\bar{x}_{\ell-1}^{(l)}\right), \qquad (33)$$

where $\bar{x}_{\ell-1}^{(l)}$ is the entry-wise projection of the previous estimation $\widehat{x}_{\ell-1}^{(l)}$ onto signal constellation points. The output of the MMSE-PIC detector is $(\bar{x}_L^{(0)}, \bar{x}_L^{(1)}, \ldots, \bar{x}_L^{(N-1)})$. Note that the MMSE-PIC detector has no trainable parameters and thus the training process is not necessary.

B. DetNet-PIC

Next, we give a brief review of the DetNet-PIC used for the simulations presented in Section IV. The derivation of the DetNet follows [19], [20], and the DetNet-PIC for OAM-MIMO detection is derived by integrating a PIC function with a mode-wise DetNet in a manner similar to the MMSE-PIC and the ADMM-PIC. The mode-wise DetNet consists of L layers, each of which performs the following procedure [20]:

$$z_\ell^{(k)} = x_\ell^{(k)} - \delta_{1,\ell}^{(k)} H^{(k,k)T} y^{(k)} + \delta_{2,\ell}^{(k)} H^{(k,k)T} H^{(k,k)} x_\ell^{(k)} \quad (34)$$

$$u_\ell^{(k)} = R\big(W_{1,\ell}^{(k)}(z_\ell^{(k)T}, v_\ell^{(k)T})^T + b_{1,\ell}^{(k)}\big) \qquad (35)$$

$$v_{\ell+1}^{(k)} = W_{2,\ell}^{(k)} u_\ell^{(k)} + b_{2,\ell}^{(k)} \qquad (36)$$

$$x_{\ell+1}^{(k)} = W_{3,\ell}^{(k)} u_\ell^{(k)} + b_{3,\ell}^{(k)} \qquad (37)$$

where $R(x)$ denotes the ReLU function, and where $\boldsymbol{u}_\ell^{(k)}$ and $\boldsymbol{v}_\ell^{(k)}$ are real intermediate internal vectors. The setting of the dimensions of the intermediate internal vectors is heuristic and depends on the modulation scheme used [20]. In the simulations of Section IV, the dimensions of $\boldsymbol{u}_\ell^{(k)}$ and $\boldsymbol{v}_\ell^{(k)}$ were set to $32M$ and $4M$, respectively. Thus the learnable parameters in the ℓ-th layer were $\delta_{1,\ell}^{(k)}, \delta_{2,\ell}^{(k)} \in \mathbb{R}$, $\boldsymbol{W}_{1,\ell}^{(k)} \in \mathbb{R}^{32M \times 6M}$, $\boldsymbol{b}_{1,\ell}^{(k)} \in \mathbb{R}^{32M}$, $\boldsymbol{W}_{2,\ell}^{(k)} \in \mathbb{R}^{4M \times 32M}$, $\boldsymbol{b}_{2,\ell}^{(k)} \in \mathbb{R}^{4M}$, $\boldsymbol{W}_{3,\ell}^{(k)} \in \mathbb{R}^{2M \times 32M}$, and $\boldsymbol{b}_{3,\ell}^{(k)} \in \mathbb{R}^{2M}$. The DetNet-PIC can be obtained by replacing $\boldsymbol{y}^{(k)}$ in (34) with $\widetilde{\boldsymbol{y}}_\ell^{(k)}$ given as

$$\widetilde{\boldsymbol{y}}_\ell^{(k)} = \boldsymbol{y}^{(k)} - \boldsymbol{\gamma}_\ell^{(k)} \circ \sum_{l=0, l\neq k}^{N-1} \boldsymbol{H}^{(k,l)} \boldsymbol{x}_\ell^{(l)}, \qquad (38)$$

where $\boldsymbol{\gamma}_\ell^{(k)}$ is a learnable parameter vector of dimension $2M$. Thus, the total number of learnable parameters per each layer becomes $(384M^2 + 38M + 2)N + 2MN$, which is much larger than $8MN$ (the number of learnable parameters per each layer in Algorithm 1).

ACKNOWLEDGMENT

REFERENCES

[1] P. F. Driessen and G. J. Foschini, "On the capacity formula for multiple input-multiple output wireless channels: A geometric interpretation," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 173–176, Feb. 1999.

[2] L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman, "Orbital angular momentum of light and the transformation of laguerre-Gaussian laser modes," *Phys. Rev. A, Gen. Phys.*, vol. 45, no. 11, pp. 8185–8189, Jun. 1992.

[3] F. Tamburini, E. Mari, A. Sponselli, B. Thidé, A. Bianchini, and F. Romanato, "Encoding many channels on the same frequency through radio vorticity: First experimental test," *New J. Phys.*, vol. 14, no. 3, Mar. 2012, Art. no. 033001.

[4] H. Sasaki, D. Lee, H. Fukumoto, Y. Yagi, T. Kaho, H. Shiba, and T. Shimizu, "Experiment on Over-100-Gbps wireless transmission with OAM-MIMO multiplexing system in 28-GHz band," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, UAE, Dec. 2018, pp. 1–6.

[5] Y. Yagi, H. Sasaki, T. Yamada, and D. Lee, "200 Gbit/s wireless transmission using dual-polarized OAM-MIMO multiplexing on 28 GHz band," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–4.

[6] H. Sasaki, Y. Yagi, T. Yamada, and D. Lee, "Field experimental demonstration on OAM-MIMO wireless transmission on 28 GHz band," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–4.

[7] M. Hirabe, R. Zenkyu, H. Miyamoto, and K. Ikuta and, E Sasaki, "40 M transmission of OAM mode and polarization multiplexing in E-band," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[8] Y. Yan, G. Xie, M. P. J. Lavery, H. Huang, N. Ahmed, C. Bao, Y. Ren, Y. Cao, L. Li, Z. Zhao, A. F. Molisch, M. Tur, M. J. Padgett, and A. E. Willner, "High-capacity millimetre-wave communications with orbital angular momentum multiplexing," *Nature Commun.*, vol. 5, no. 1, p. 4876, Sep. 2014.

[9] F. E. Mahmouli and S. D. Walker, "4-gbps uncompressed video transmission over a 60-GHz orbital angular momentum wireless channel," *IEEE Wireless Commun. Lett.*, vol. 2, no. 2, pp. 223–226, Apr. 2013.

[10] Y. Ren, L. Li, G. Xie, Y. Yan, Y. Cao, H. Huang, N. Ahmed, Z. Zhao, P. Liao, C. Zhang, G. Caire, A. F. Molisch, M. Tur, and A. E. Willner, "Line-of-Sight millimeter-wave communications using orbital angular momentum multiplexing combined with conventional spatial multiplexing," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3151–3161, May 2017.

[11] Z. Zhao, Y. Ren, G. Xie, Y. Yan, L. Li, H. Huang, C. Bao, N. Ahmed, M. P. Lavery, C. Zhang, N. Ashrafi, S. Ashrafi, S. Talwar, S. Sajuyigbe, M. Tur, A. F. Molisch, and A. E. Willner, "Experimental demonstration of 16-Gbit/s millimeter-wave communications link using thin metamaterial plates to generate data-carrying orbital-angular-momentum beams," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 1392–1397.

[12] W. Zhang, S. Zheng, X. Hui, R. Dong, X. Jin, H. Chi, and X. Zhang, "Mode division multiplexing communication using microwave orbital angular momentum: An experimental study," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 1308–1318, Feb. 2017.

[13] O. Edfors and A. J. Johansson, "Is orbital angular momentum (OAM) based radio communication an unexploited area?" *IEEE Trans. Antennas Propag.*, vol. 60, no. 2, pp. 1126–1131, Feb. 2012.

[14] K. A. Opare and Y. Kuang, "Performance of an ideal wireless orbital angular momentum communication system using multiple-input multiple-output techniques," in *Proc. Int. Conf. Telecommun. Multimedia (TEMU)*, Jul. 2014, pp. 144–149.

[15] F. Bohagen, P. Orten, and G. E. Oien, "Design of optimal high-rank Line-of-Sight MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1420–1425, Apr. 2007.

[16] D. Shin, E. Park, J. Kang, J. Myung, and J. Kang, "Identification of non-ideal receiver condition for orbital angular momentum transmission," in *Proc. IEEE 79th Veh. Technol. Conf. (VTC Spring)*, Seoul, South Korea, May 2014, pp. 1–5.

[17] Y. Yuan, Z. Zhang, J. Cang, H. Wu, and C. Zhong, "Capacity analysis of UCA-based OAM multiplexing communication system," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Nanjing, China, Oct. 2015, pp. 1–5.

[18] R. Chen, H. Xu, J. Li, and Y. Zhang, "Misalignment-robust receiving scheme for UCA-based OAM communication systems," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Sydney, NSW, Australia, Jun. 2017, pp. 1–5.

[19] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Sapporo, Japan, Jul. 2017, pp. 1–5.

[20] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2554–2564, May 2019.

[21] M.-W. Un, M. Shao, and W.-K. C. Ma andP Ching, "Deep MIMO detection using ADMM unfolding," in *Proc. IEEE Data Sci. Workshop (DSW)*, Jun. 2019, pp. 333–337.

[22] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[23] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 10–18.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Intern. Conf. Learn. Represent. (ICLR)*, San Diego, CA, pp. 1–15, 2015.

[25] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Conf. Operating Syst. Design Implement. (OSDI)*, Berkeley, CA, USA, 2016, pp. 265–283.

NORIFUMI KAMIYA (Member, IEEE) received the B.E. degree from Yokohama National University, in 1990, and the Ph.D. degree from The University of Tokyo, in 2000, all in electrical engineering. From 2001 to 2002, he was a Visiting Researcher with the University of Hawaii at Manoa. He is currently a Senior Principle Researcher with the System Platform Research Laboratories, NEC Corporation, Japan. His current research interests include statistical signal processing and coding for wireless and optical communication systems.

$\bullet \bullet \bullet$