# Detection for Hybrid Beamforming Millimeter Wave Massive MIMO Systems

Kazem Izadinasab, *Student Member, IEEE*, Ahmed Wagdy Shaban, *Student Member, IEEE*, and Oussama Damen, *Senior Member, IEEE*

***Abstract*— In this letter, we improve the error performance of hybrid beamforming millimeter wave (mmWave) massive multi-input multi-output (MIMO) systems by designing detectors for such systems. First, we discuss the effect of the mmWave channel parameters and hybrid beamforming settings on the equivalent channel, which consists of the precoder, mmWave channel, and combiner. Then, we propose a low-complexity near-optimal signal detection scheme for the equivalent channel. Using computer simulations, it is shown that the error performance can be significantly improved and computational complexity reduced compared to those of state-of-the-art MIMO detection schemes. We achieve this reduced complexity by exploiting the available structure in the equivalent channel.**

***Index Terms*— Massive MIMO, maximum likelihood detection, subspace detection, hybrid beamforming.**

## I. Introduction

HYBRID beamforming is considered to be the viable realization of massive multi-input multi-output (MIMO) systems in the millimeter wave (mmWave) bands. Its viability lies in the low implementation and operational costs of utilizing fewer radio frequency (RF) chains at base stations and/or in users' equipment. However, this approach comes at the price of some performance loss. Existing beamforming designs aim at minimizing this performance loss under perfect channel state information (CSI) [1]–[3], partial CSI [4] and limited feedback CSI [5] at transmitters assumptions.

The majority of hybrid beamforming designs under different CSI frameworks are based on the approach of approximating the optimal fully digital beamformer. The *closeness* of the hybrid beamformers to the optimal fully digital one is heuristically measured, in all these existing designs, by the Euclidean distance. Under some strict assumptions, minimizing the Euclidean distance results in minimizing the mutual information loss due to hybrid beamforming. The effectiveness of the Euclidean distance for approximating the optimal beamformer in terms of spectral efficiency does not directly imply its efficacy in terms of more-practical metrics. One reason is that the spectral efficiency performance theoretically depends on the largest eigenvalues of the channel, whereas the error performance is practically dictated by the stream with the lowest signal-to-noise ratio (SNR). Therefore, the approximation error can aggravate the error performance.

Manuscript received November 21, 2020; revised December 15, 2020; accepted December 20, 2020. Date of publication December 29, 2020; date of current version April 9, 2021. This work was supported by an NSERC Collaborative Research and Development grant. The associate editor coordinating the review of this letter and approving it for publication was Q. Wu. *(Corresponding author: Kazem Izadinasab.)*

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: mkizadinasab@uwaterloo.ca; ahmed.w.shaban@uwaterloo.ca; mdamen@uwaterloo.ca).

Digital Object Identifier 10.1109/LCOMM.2020.3047994

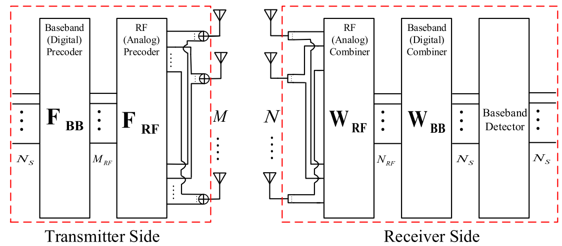



Fig. 1. System model of hybrid beamforming and detection for mmWave massive MIMO systems.

In this work, we improve the error performance of hybrid beamforming mmWave massive MIMO systems. We consider a signal detection scheme using the equivalent channel consisting of the precoder, mmWave channel, and combiner. Our proposed detector achieves an error performance close to that of the optimal maximum likelihood (ML) with significantly reduced computational complexity. In [6], the sphere decoder (SD) is proposed for mmWave massive MIMO systems by considering the diagonal matrix of the singular values as the effective channel. However, our equivalent channel matrix considers the effect of off-diagonal elements on the detection process. Moreover, the SD suffers from varying complexity with an exponential average complexity, whereas our proposed detector has a fixed and polynomial complexity and can be implemented in a fully parallel manner. We note that our proposed scheme can also be applied for intelligent reflecting surface (IRS)-aided hybrid beamforming frameworks [7].

## II. System Model and Related Works

### A. System Model

We consider a single-user MIMO system with large antenna arrays and a limited number of RF chains in the mmWave environment where the transmitter and receiver are, respectively, equipped with $M$ and $N$ antenna array elements as shown in Fig. 1. The transmitter sends $N_s$ independent data streams to the receiver with the help of $M_{\mathrm{RF}}$ RF chains. The entries of the data vector $\mathbf{s}$, which is of size $N_s \times 1$, are chosen from a given constellation $\mathcal{X}$, where $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \frac{1}{N_s}\mathbf{I}_{N_s}$. The data vector is processed by a hybrid beamforming technique in which a digital (baseband) precoder $\mathbf{F}_{\mathrm{BB}}$ of size $M_{\mathrm{RF}} \times N_s$ followed by an analog (RF) precoder $\mathbf{F}_{\mathrm{RF}}$ of size $M \times M_{\mathrm{RF}}$ are applied. Therefore, the vector $\mathbf{x} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s}$ is sent over the channel where the total transmitted power is normalized such that $\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F^2 = N_s$. The received vector of size $N \times 1$ is given by

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\rho$ is the average received power and $\mathbf{n}$ is the white Gaussian noise vector with $\mathcal{CN}(0, N_0)$ i.i.d. entries. $\mathbf{H}$ is the fading channel matrix, which is widely modeled in the mmWave literature by the clustered model [1]–[3]

$$\mathbf{H} = \sqrt{\frac{MN}{N_{\text{cl}}N_{\text{ray}}}} \sum_{i=1}^{N_{\text{cl}}} \sum_{j=1}^{N_{\text{ray}}} \alpha_{il}\mathbf{a}_{\text{r}}(\phi_{il}^{\text{r}}, \theta_{il}^{\text{r}})\mathbf{a}_{\text{t}}(\phi_{il}^{\text{t}}, \theta_{il}^{\text{t}})^H, \tag{2}$$

where $N_{\text{cl}}$ is the number of clusters, and $N_{\text{ray}}$ is the number of contributing rays in each cluster. Hence, the total number of paths is $L = N_{\text{cl}}N_{\text{ray}}$. Moreover, $\alpha_{il}$ is the complex gain of the $j$-th ray in the $i$-th cluster, and $\mathbf{a}_{\text{t}}(\phi_{il}^{\text{t}}, \theta_{il}^{\text{t}})$ is the transmit antenna array response vector of length $M$ for given azimuth and elevation angles of departure, respectively, denoted by $\phi_{il}^{t}$ and $\theta_{il}^{t}$. Similarly, $\mathbf{a}_{\text{r}}(\phi_{il}^{\text{r}}, \theta_{il}^{\text{r}})$ is the receive antenna array response vector of length $N$ for given azimuth and elevation angles of arrival. We consider uniform planar arrays (UPAs) at the transmitter and receiver. We assume that the received signal is processed by $\mathbf{W}_{\text{RF}}$, i.e., the analog combiner, then by $\mathbf{W}_{\text{BB}}$, the digital precoder, and thereby the output of combiners is

$$\tilde{\mathbf{y}} = \sqrt{\rho}\mathbf{W}_{\text{BB}}^H\mathbf{W}_{\text{RF}}^H\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{s} + \mathbf{W}_{\text{BB}}^H\mathbf{W}_{\text{RF}}^H\mathbf{n}. \tag{3}$$

After that, the signal combining stage is followed by a detection stage that estimates the transmitted data vector.

### *B. Prior Works*

The optimal fully digital precoder and combiner under a mutual information performance metric is given by [8]

$$\mathbf{F}_{\text{opt}} = \bar{\mathbf{V}}_{N_s}, \mathbf{W}_{\text{opt}} = \bar{\mathbf{U}}_{N_s}, \tag{4}$$

where $\bar{\mathbf{V}}_{N_s}$ and $\bar{\mathbf{U}}_{N_s}$ are the largest $N_s$ right and left singular vectors of the channel matrix, respectively, obtained by the singular value decomposition (SVD) of $\mathbf{H}$, i.e., $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$. Most prior works on hybrid beamforming adopted the approach of decomposing the optimal beamforming (precoding/combining) matrix into two beamforming matrices, one analog and the other digital [1]–[4] using the following approximation

$$\begin{aligned} &\underset{\mathbf{F}_{\text{RF}},\mathbf{F}_{\text{BB}}}{\operatorname{argmin}} \parallel \mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}} \parallel_F^2 \\ &\text{subject to } \mathbf{F}_{\text{RF}} \in \mathbb{U}^{M\times M_{\text{RF}}}, \\ &\qquad\qquad \parallel \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}} \parallel_F^2 = N_s, \end{aligned} \tag{5}$$

where $\mathbb{U}^{M\times M_{\text{RF}}}$ is the space of $M \times M_{\text{RF}}$ matrices that have unit magnitude entries because the analog precoder is implemented by constant magnitude phase shifters. The same problem formulation is applied to the receiver side for approximating $\mathbf{W}_{\text{opt}}$ as $\mathbf{W}_{\text{RF}}\mathbf{W}_{\text{BB}}$ by using $N_{\text{RF}}$ RF chains. We note that there is no closed-form solution for this non-convex optimization problem, and all the prior works offer sub-optimal solutions ([1]–[3] and references therein). Although it is shown [9] that optimizing mutual information and bit error rate (BER) requires minimizing the chordal and projection two-norm subspace distances, respectively, the majority of prior works utilize the Euclidean distance instead of these sophisticated distances. This use is advised because of the high computational complexity incurred by such sophisticated distances compared to the Euclidean distance, and since the Euclidean distance yields a good approximation of the chordal distance. However, the Euclidean distance is not necessarily the best metric to use in measuring the closeness of the hybrid beamformers to the optimal one so as to minimize the BER. Using the Euclidean distance directly results in high errors when the optimal beamformer is approximated using hybrid ones. The result is inter-symbol interference (ISI) affecting the stream with the lowest SNR and aggravating the total BER performance, as will be shown using computer simulations in Section V. To address this problem, we propose adding a low-complexity detection stage after the hybrid beamforming (precoding/combining) stage in order to enhance the BER performance of the system.

## III. Proposed Signal Detection Algorithm

### *A. Equivalent Channel Matrix*

We consider the equivalent channel consisting of the precoder, mmWave channel, and combiner, i.e.,

$$\mathbf{H}_{\text{eq}} = \mathbf{W}^H\mathbf{H}\mathbf{F} \tag{6}$$

for the detection process. Hence, we have

$$\tilde{\mathbf{y}} = \sqrt{\rho}\mathbf{H}_{\text{eq}}\mathbf{s} + \mathbf{n}', \tag{7}$$

where $\mathbf{n}' = \mathbf{W}^H\mathbf{n}$. When the optimal fully digital SVD-based precoder and combiner in (4) are used, this equivalent channel will be $\mathbf{\Sigma}_1$, the $N_s \times N_s$ partition of $\mathbf{\Sigma}$, and the optimal detector will be a simple symbol-by-symbol demodulation. However, for the hybrid beamforming structure, there is a mismatch between the optimal precoder/combiner and the hybrid precoder/combiner due to the approximations that makes the equivalent channel non-diagonal. The non-zero off-diagonal entries are problematic for signal detection, especially for the symbols corresponding to the last columns of the equivalent channel, because the diagonal entries are in a descending order as a result of the SVD operation. The quality of the approximation (small off-diagonal entries) for each hybrid beamforming scheme affects the error gap between a simple detector and the optimal ML solution.

In practical mmWave massive MIMO systems, the number of data streams, i.e., $N_s$ is as high as 8 [10]. Because of the near-diagonal structure of the equivalent channel in (6), it is expected that MIMO detection techniques such as the fixed-complexity sphere decoder (FSD) [11] and the subspace detection scheme in [12] can be extensively simplified by exploiting the structure.

### *B. Proposed Low-Complexity Near-Optimal Detector*

The proposed detector consists of the following steps:

*Channel Ordering*: We consider the regularized (or augmented) channel matrix, i.e.,

$$\tilde{\mathbf{H}}_{\text{eq}} = \begin{bmatrix} \mathbf{H}_{\text{eq}} \\ \sqrt{\alpha}\,\mathbf{I}_{N_s} \end{bmatrix}, \tag{8}$$

**Algorithm 1** Multiple MMSE-SIC Subspace Detector

**Inputs**: $\mathbf{H}_{\text{eq}} \in \mathbb{C}^{N_s \times N_s}$, $\mathbf{y} \in \mathbb{C}^{N_s \times 1}$, $\mathcal{X}$.
**Parameters**: $1 \le c \le N_s$.

1: Channel Ordering:
   1) Find $c$ worst columns of matrix $\mathbf{H}_{\text{eq}}$.
   2) If $c = 1$, keep this worst column at the end of the ordered matrix denoted as $\mathbf{H}_{\text{O}}$. Otherwise, i.e., if $c > 1$, put these $c$ worst columns at the $c$ first column of the ordered matrix.
   3) Place the $N_s - c$ best remaining columns in order at the remaining columns of $\mathbf{H}_{\text{O}}$.
2: Perform QRD on $\tilde{\mathbf{H}}_{\text{O}}$; $\tilde{\mathbf{H}}_{\text{O}} = \tilde{\mathbf{Q}}\mathbf{R}$.
3: $\mathbf{y}' = \mathbf{Q}^H \mathbf{y} = \mathbf{R}\mathbf{s} + \mathbf{w}$; $\mathbf{Q}$ is the $N_s \times N_s$ part of $\tilde{\mathbf{Q}}$.
4: Perform the subspace detection on $\mathbf{y}'$:
5: For $i' = 1, \ldots, c$:
   1) If $c = 1$, $i = N_s$.
   2) If $c > 1$, $i = i'$.
   3) $\mathbf{y}' = \mathbf{R}_{[i]}\mathbf{s}_{[i]} + \mathbf{r}_i s_i + \mathbf{w}$.
   4) For $j = 1, \ldots, Q = |\mathcal{X}|$:
      a) Fix $\hat{s}_i^{(j)} = a_j$, $a_j \in \mathcal{X}$.
      b) If $c = 1$, $\mathbf{z}^{[i]} = \mathbf{y}' - \mathbf{r}_i s_i$.
      c) If $c > 1$, perform QRD on $\mathbf{R}_{[i]}$; i.e., $\mathbf{R}_{[i]} = \mathbf{Q}^{[i]}\mathbf{R}^{[i]}$. Then, $\mathbf{z}^{[i]} = \mathbf{Q}^{[i]H}(\mathbf{y}' - \mathbf{r}_i s_i) = \mathbf{R}^{[i]}\mathbf{s}_{[i]} + \mathbf{Q}^{[i]H}\mathbf{w}$.
      a) Detect $\hat{\mathbf{s}}_{[i]}^{(j)}$ by performing SIC on $\mathbf{z}^{[i]}$, and slice it to the original constellation points.
      b) Calculate $d_k = \|\mathbf{y}' - \mathbf{R}_{[i]}\hat{\mathbf{s}}_{[i]}^{(j)} - \mathbf{r}_i \hat{s}_i^{(j)}\|^2$; $k = (i'-1)|\mathcal{X}| + j$.
6: Find $d_{min} = \min\limits_{k=1,..,cQ} d_k$ and its corresponding $(i^\star, i'^\star, j^\star)$. Declare $\hat{\mathbf{s}} = [\hat{\mathbf{s}}_{[i^\star]}^{(j^\star)}, \hat{s}_{i^\star}^{(j^\star)}]$ to be detected symbols after an appropriate reordering.

**Output**: $\hat{\mathbf{s}}$

where $\alpha = \frac{N_0}{\rho}$. Depending on the value of $c$, which is the number of columns that are considered for subspace detection step in (11), an appropriate channel ordering is performed, and the ordered matrix is denoted as $\mathbf{H}_{\text{O}}$.

*MMSE Filtering*: We perform the QR decomposition (QRD) on $\mathbf{H}_{\text{O}}$, i.e.,

$$\tilde{\mathbf{H}}_{\text{eq}} = \tilde{\mathbf{Q}}\mathbf{R}, \tag{9}$$

where $\tilde{\mathbf{Q}}$ is an $2N_s \times N_s$ matrix with orthonormal columns, and $\mathbf{R}$ is an $N_s \times N_s$ upper-triangular matrix with positive diagonal entries. By considering $\mathbf{Q}$ as the first $N_s$ rows of $\tilde{\mathbf{Q}}$ one writes

$$\mathbf{y}' = \mathbf{Q}^H \mathbf{y} = \mathbf{R}\mathbf{s} + \mathbf{w}, \tag{10}$$

where $\mathbf{w} = \mathbf{Q}^H\mathbf{n}' - [\mathbf{R} - \mathbf{Q}^H\mathbf{H}_{\text{eq}}]\mathbf{s}$.

*Subspace Detection*: The idea of subspace detection is to divide the symbol vector into two subvectors. Here, we consider a subvector of size one for the detection of one symbol corresponding to one column. Hence, we have

$$\mathbf{y}' = \mathbf{R}_{[i]}\mathbf{s}_{[i]} + \mathbf{r}_i s_i + \mathbf{w}, \tag{11}$$

where $\mathbf{R}_{[i]}$ is the submatrix of $\mathbf{R}$ obtained by removing the $i$-th column. The constellation points are searched to find $s_i$. When $c > 1$, we examine $c$ worst columns in a round-robin fashion for subspace detection.

*Successive Interference Cancellation (SIC)*: After removing the affect of each considered constellation point, SIC is performed to detect the remaining symbols. If $c = 1$, $\mathbf{r}_{N_s}$ is used for subspace detection. Hence, SIC is performed using the upper-triangular structure of $\mathbf{R}_{[N_s]}$. However, when $c > 1$, a second QRD is required in order for $\mathbf{R}_{[i]}$ to have an upper-triangular form for SIC. The efficacy of the proposed channel ordering for SIC is two-fold: first, by considering $\mathbf{r}_i$ for subspace detection, the symbols corresponding to the remaining $c-1$ worst columns are detected at the last steps of SIC, thereby improving the BER. Second, the remaining submatrix $\mathbf{R}_{[i]}$ can be converted into an upper-triangular form using *Givens rotations* technique [13] as we need up to $N_s - 1$ rotations. For example, for

$$\mathbf{R}_{[1]} = \begin{bmatrix} r_{12} & \ldots & r_{1N_s} \\ r_{22} & \ldots & r_{2N_s} \\ \vdots & \ddots & \vdots \\ 0 & \ldots & r_{N_sN_s} \end{bmatrix},$$

we only need to remove the entries of the subdiagonal, i.e., $\{r_{22}, \ldots, r_{N_sN_s}\}$. Each rotation changes two rows of the submatrix and needs up to $4(N_s - 1)$ multiplications and $2(N_s - 1)$ additions. As a result, $\mathcal{O}(cN_s^2)$ operations are required for the second QRD of all $c$ submatrices.

*Euclidean Distance Test*: among all detected vectors of transmitted symbols, the one with the minimum Euclidean distance is declared the detected vector.

Changing the parameters of Algorithm 1 converts it to other detection schemes with reduced complexity. For example, when $c = 1$, it becomes equivalent to the FSD scheme with one full-search level [11]. For $c = N_s$, other than the ordering, it is similar to the multiple subspace detection method in [12]. Also, if the subspace detection is omitted, the algorithm becomes equivalent to the MMSE-SIC scheme but with the addition of V-BLAST ordering [14]. Consequently, by appropriate assignment of the parameters, a detection scheme can be applied to achieve a near-optimal error performance while keeping the computational complexity as low as possible.

## IV. Computational Complexity

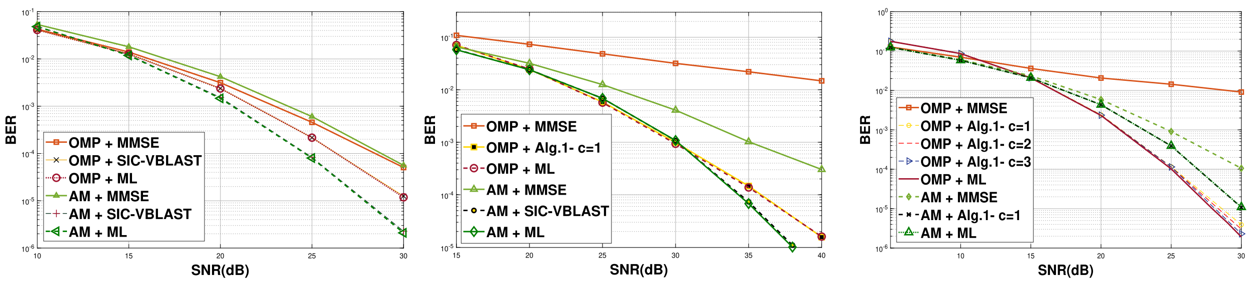
In Algorithm 1, QRD and the channel ordering both require $\mathcal{O}(N_s^3)$ operations. For multiple subspace detection, additional $\mathcal{O}(cN_s^2)$ operations are required for all second QRDs. Table I shows the detailed complexity of the proposed algorithm in terms of number of floating-point operations (FLOPs) assuming six and two FLOPs per complex multiplication and addition, respectively. The complexity of the proposed algorithm is $\mathcal{O}(N_s^3|\mathcal{X}|)$.

The computational complexity of orthogonal matching pursuit (OMP) hybrid beamforming technique [1] or its variants is $\mathcal{O}(M^2 N_{\text{RF}} N_s)$. The alternating minimization (AM) hybrid beamforming technique requires $\mathcal{O}(MN_{\text{RF}}^2\ N_s)$ operations, mainly due to partial SVD computations using rank revealing

TABLE I

COMPUTATIONAL COMPLEXITY IN TERMS OF FLOPs

<table>
<tr><th rowspan="2">Detector</th><th rowspan="2" colspan="2">Operation</th><th rowspan="2">Number of FLOPs</th><th colspan="2">$N_S = 8$</th></tr>
<tr><th colspan="2">64-QAM</th></tr>
<tr><td rowspan="5">FSD</td><td rowspan="2">Preprocessing:</td><td>FSD ordering+ QRD [14]</td><td>$28N_S^3 + \frac{113}{3}N_S^2 + 24N_S - 60$</td><td rowspan="2" colspan="2">$475,630$<br>$(p = 2)$</td></tr>
<tr><td>$\mathbf{R}_2\mathbf{x}_2; \mathbf{x}_2 \in \mathcal{X}^p$ for $p$ worst column of $\mathbf{R}$</td><td>$(8N_S p - 2N_S)|\mathcal{X}|^p$</td></tr>
<tr><td rowspan="3">Detection:</td><td>$\mathbf{y}' = \mathbf{Q}^H \mathbf{y}$</td><td>$(8N_S - 2)N_S$</td><td rowspan="3" colspan="2">$1,647,088$<br>$(p = 2)$</td></tr>
<tr><td>SIC</td><td>$(4(N_S - p)(N_S - p + 1) + 2p)|\mathcal{X}|^p$</td></tr>
<tr><td>Euclidean Distance Test</td><td>$(4N_S^2 + 4N_S - 2)|\mathcal{X}|^p$</td></tr>
<tr><td rowspan="7">Multiple MMSE-SIC Subspace Detector (Algorithm 1)</td><td rowspan="3">Preprocessing:</td><td>Ordering + QRD [14]</td><td>$28N_S^3 + \frac{113}{3}N_S^2 + 24N_S - 60$</td><td rowspan="3">$19,951$<br>$(c = 1)$</td><td rowspan="3">$26,935$<br>$(c = 3)$</td></tr>
<tr><td>QRD of $\mathbf{R}^{[i]}$; $i = 1, \ldots, c$ (if $c > 1$)</td><td>$c(6N_S^2 - 14N_S + 8)$</td></tr>
<tr><td>$\mathbf{r}_i a_j$; $i = 1, \ldots, c$; $j = 1, \ldots, |\mathcal{X}|$</td><td>$6N_S c|\mathcal{X}|$</td></tr>
<tr><td rowspan="4">Detection:</td><td>$\mathbf{y}' = \mathbf{Q}^H \mathbf{y}$</td><td>$8N_S^2 - 2N_S$</td><td rowspan="4">$33,264$<br>$(c = 1)$</td><td rowspan="4">$100,288$<br>$(c = 3)$</td></tr>
<tr><td>$\mathbf{z}^{[i]} = \mathbf{Q}^{[i]H}(\mathbf{y}' - \mathbf{r}_i a_j)$; $i = 1, \ldots, c$ (if $c > 1$)</td><td>$(8N_S^2 - 2N_S)c$</td></tr>
<tr><td>SIC</td><td>$(4N_S^2 - 4N_S + 2)c|\mathcal{X}|$</td></tr>
<tr><td>Euclidean Distance Test</td><td>$(4N_S^2 + 4N_S - 2)c|\mathcal{X}|$</td></tr>
</table>




(a) $N_s = 4$, $N_{\mathrm{RF}} = 6$, $N_{\mathrm{cl}} = 3$, $N_{\mathrm{ray}} = 4$, $L = 12$, 256QAM (b) $N_s = 6$, $N_{\mathrm{RF}} = 6$, $N_{\mathrm{cl}} = 3$, $N_{\mathrm{ray}} = 4$, $L = 12$, 256QAM (c) $N_s = 8$, $N_{\mathrm{RF}} = 8$, $N_{\mathrm{cl}} = 5$, $N_{\mathrm{ray}} = 6$, $L = 30$, 64-QAM

Fig. 2. BER of hybrid beamforming with detection for a $64 \times 16$ UPA mmWave massive MIMO System.

QRD [15]. One has to add the computational complexity of calculating the optimal beamformers, which is also dominated by the complexity of partial SVD computation. This computation requires $\mathcal{O}(MNN_s)$ operations when using rank revealing QRD to calculate $\mathbf{U}$ and $\mathbf{V}$ [15]. Hence, the overall computational complexity of the hybrid beamforming processes is $\mathcal{O}(M(N + N_{\mathrm{RF}}^2)N_s)$.

## V. SIMULATION RESULTS

In this section, we investigate the error performance of the proposed scheme for both OMP and AM hybrid beamforming techniques. In all simulations, we perform hybrid beamforming for both precoding and combining by assuming $M_{\mathrm{RF}} = N_{\mathrm{RF}}$. The OMP algorithm has a fixed number of iterations equal to the number of RF chains. Hence, for a fair comparison, we set the number of iterations of the AM technique to $N_{\mathrm{RF}}$. We consider i.i.d. azimuth angles of departure and arrival with Laplacian distribution with means uniformly distributed between $0$ and $2\pi$, and the angular spreads are $7.5°$ [1]. We compare our proposed scheme with the linear MMSE receiver for minimizing error performance [1], [8]. This MMSE receiver is equivalent to our system model when $\mathbf{W}_{\mathrm{opt}} = \bar{\mathbf{U}}_{N_s}$ is used for the combiner followed by the MMSE detector.

Fig. 2a shows the error performance for $N_s = 4$, $N_{\mathrm{RF}} = 6$ and $L = 12$. In this figure, the MMSE-SIC detector with the optimal V-BLAST ordering achieves a performance close to that of the ML for both AM and OMP hybrid algorithms. Note that the MMSE-SIC with V-BLAST can be implemented by skipping the subspace detection in Algorithm 1. By using detection, gains of about 2-dB and 4-dB are achieved, respectively, for OMP and AM at the BER of $10^{-4}$.

In Fig. 2b, $N_s = N_{\mathrm{RF}} = 6$ and $L = 12$. The proposed algorithm significantly improves the error performance for both OMP and AM techniques. For AM, the MMSE-SIC with V-BLAST suffices to achieve the ML performance, while for OMP, Algorithm 1 with $c = 1$ is required. In Fig. 2c, we increase the number of streams such that $N_s = N_{\mathrm{RF}} = 8$. The proposed algorithm with $c = 1$ improves the error performance for both OMP and AM techniques. $c = 1$ is sufficient for AM technique to achieve ML error performance. For OMP, however, with $c = 1$, about 1-dB gap appears at high SNRs. This gap is removed by considering $c = 3$. As Table I shows, for this system, the computational complexity of the proposed algorithm is significantly reduced even for $c = 3$ compared to that of the FSD scheme. One may further reduce complexity by considering a smaller search size over the given constellation [16], i.e., $Q$ in Algorithm 1.

From the simulation results, we observe that the OMP technique performs poorly when $N_{\mathrm{RF}} = N_s$, or when $L$ is large. Moreover, the error performance of AM technique is highly susceptible to the parameter $L$ such that, for a large $L$, the gap between the MMSE and the ML becomes smaller even for the edge-case $N_s = N_{\mathrm{RF}}$, whereas for a small $L$, there will be a performance gap.

## VI. CONCLUSION

The approximation error introduced to mmWave massive MIMO systems due to hybrid beamforming techniques can significantly degrade the error performance of these systems.

Consequently, we have proposed a signal detection algorithm that can improve the error performance of such systems. Our proposed algorithm approaches the error performance of the optimal ML detector, while also reducing the computational complexity compared to the ML or other conventional detectors.

## REFERENCES

[1] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[2] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[3] R. Rajashekar and L. Hanzo, "Iterative matrix decomposition aided block diagonalization for mm-wave multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1372–1384, Mar. 2017.

[4] J. Jin, C. Xiao, W. Chen, and Y. Wu, "Channel-statistics-based hybrid precoding for millimeter-wave MIMO systems with dynamic subarrays," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 3991–4003, Jun. 2019.

[5] A. W. Shaban, O. Damen, Y. Xin, and E. Au, "Statistically-aided codebook-based hybrid precoding for mmWave single-user MIMO systems," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6.

[6] M. Alouzi, F. Chan, and C. D'Amours, "Sphere decoding for millimeter wave massive MIMO," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2019, pp. 1–6.

[7] B. Ning, Z. Chen, W. Chen, Y. Du, and J. Fang, "Channel estimation and hybrid beamforming for reconfigurable intelligent surfaces assisted THz communications," 2019, *arXiv:1912.11662*. [Online]. Available: https://arxiv.org/abs/1912.11662

[8] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[9] D. J. Love and R. W. Heath, Jr., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[10] *New Radio; Physical Channels and Modulation*, Standard TS 38.211 V16.0.0, Dec. 2019.

[11] L. Barbero and J. Thompson, "Fixing the complexity of the sphere decoder for MIMO detection," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2131–2142, Jun. 2008.

[12] M. K. Izadinasab and O. Damen, "Bridging the gap between MMSE-DFE and optimal detection of MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 220–231, Jan. 2020.

[13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.

[14] D. Wubben, R. Bohnke, V. Kuhn, and K.-D. Kammeyer, "MMSE extension of V-BLAST based on sorted QR decomposition," in *Proc. IEEE 58th Veh. Technol. Conf. (VTC-Fall)*, Oct. 2003, pp. 508–512.

[15] M. Gu and S. C. Eisenstat, "Efficient algorithms for computing a strong rank-revealing QR factorization," *SIAM J. Sci. Comput.*, vol. 17, no. 4, pp. 848–869, Jul. 1996.

[16] M. K. Izadinasab and O. Damen, "Reduced complexity ordering in subspace MIMO detection algorithms," in *Proc. 16th Can. Workshop Inf. Theory (CWIT)*, Jun. 2019, pp. 1–5.