Media-Based Modulation for the Uplink in Massive MIMO Systems

Bharath Shamasundar†, Swaroop Jacob‡, Lakshmi Narasimhan T*, and A. Chockalingam†
†Department of ECE, Indian Institute of Science, Bangalore 560012
‡ Presently with Cisco Systems Private Limited, Bangalore 560087
* Presently with Department of EE, Indian Institute of Technology Palakkad, Kerala 678557

***Abstract*—Media-based modulation (MBM) is an attractive modulation scheme which can achieve high data rates using multiple radio frequency (RF) mirrors (parasitic elements) and fewer transmit antennas/RF chains. In this paper, we consider MBM for the uplink in a massive MIMO system consisting of tens of users and tens to hundreds of base station (BS) receive antennas. Each user employs MBM with one transmit antenna and multiple RF mirrors placed near it. The ON/OFF status of the RF mirrors conveys information bits in addition to the bits conveyed through conventional modulation symbols. We show that the multiuser MBM (MU-MBM) system achieves very good bit error performance compared to other conventional multiuser MIMO systems. We propose two types of low-complexity detection algorithms for massive MU-MBM systems. The first type is based on compressive sensing (CS) that exploits the inherent sparse nature of the MBM transmit vectors. The second type is is based on message passing that performs both detection and channel estimation at low computational complexity. To do this, it exploits the 'channel hardening' phenomenon that occurs in large MIMO channels. It is shown that the proposed algorithms are not only computationally less complex but also provide very good bit error performance in massive MU-MBM systems.**

***Index Terms*—Media-based modulation, parasitic elements, RF mirrors, compressive sensing, structured sparsity, message passing receiver, channel hardening**

I. Introduction

Next generation of wireless communication systems are expected to deliver very high data rates with high reliability and spectral efficiency. Multiple input multiple output (MIMO) systems, where both the transmitter and receiver use multiple antennas/RF chains, is a key technology for achieving such high data rates in a power and spectral efficient manner [1]. Large-scale multiuser MIMO systems for downlink and uplink communications (e.g., massive MIMO [2]) are of interest. In multiuser MIMO downlink, a base station (BS) with large number of antennas (10s to 100s) transmit simultaneously to several users. Employing zero-forcing (ZF) precoding in the digital baseband is known to achieve near optimal performance when the number of BS antennas is large. However, ZF precoding requires the use of large number of RF chains, equal to the number of BS antennas, which is prohibitive since it increases the hardware cost and also leads to more power consumption. To overcome this hardware constraint and achieve the benefits of multiuser MIMO systems, several hybrid analog and digital precoding schemes are proposed. The main idea is to divide the precoder into a small size digital precoder requiring less number of RF chains and a large size analog precoder to achieve array gain. It has been shown that hybrid precoding can reduce the number of RF chains at the BS with negligible performance degradation compared to fully digital precoding (see [3], [4] and the references therein).

On the uplink, multiple users transmit simultaneously to a BS having multiple receive antennas. The use of complex RF hardware in the user equipment (UE) is prohibitive due to cost and operating power requirements. MIMO modulation techniques that can enable the use of fewer transmit antennas/RF chains at the UE are crucial for uplink data transmissions. Three interesting approaches that can significantly reduce the RF hardware complexity at UE are becoming popular in the recent literature. They include spatial modulation (SM) [5]-[8], load modulation (LM) [9]-[11], and media-based modulation (MBM) [12]-[16]. A key commonality in them is that they can work with only one transmit RF chain along with an array of antenna elements. The single-RF chain feature in these approaches enables RF hardware simplicity and size/cost reduction. In the following, we briefly describe these MIMO modulation schemes highlighting their relative differences and merits.

SM: An SM system uses multiple transmit antenna elements and a single RF chain at the transmitter [6]-[8]. In a given channel use, only one transmit antenna element is activated based on the information bits and a symbol from conventional modulation alphabet (say, QAM) is transmitted from the activated antenna. If n_t is the number of transmit antenna elements and $\mathbb{A}$ is the modulation alphabet used, then the achieved rate in SM is $\eta_{\text{SM}} = \lfloor \log_2 n_t \rfloor + \log_2 |\mathbb{A}|$ bits per channel use (bpcu). Therefore, SM requires an exponential increase in the number of antenna elements to increase the rate. However, limited space available in the UE prohibits the use of several transmit antenna elements, as required by SM.

LM: LM array is an interesting MIMO modulation scheme in which the load impedances of multiple transmit antenna elements (parasitic elements) are modulated according to information signals, in effect implementing the signal set in the analog domain [9]. In LM, a single power amplifier fed by a fixed voltage and frequency source drives an array of multiple transmit antenna elements. The LM implementation requires only one transmit RF chain since it uses only one power amplifier and an optional spectral shaping filter to reduce the sampling rate to symbol rate [9],[10]. The use of LM arrays with 2 antenna elements at the UEs in a multiuser MIMO system has been shown to achieve good performance

[11]. A drawback with LM is that all the transmit vectors are constrained to have the same transmit power, i.e., all the transmit signals should lie on the surface of the n_t-dimensional hypersphere. Any power mismatch can result in refection of power to the amplifier, which reduces its efficiency.

MBM: MBM is a recently proposed channel modulation scheme that can achieve high rates over multipath fading channels in rich scattering environments [12]-[16]. In MBM, digitally controlled parasitic elements, called radio frequency (RF) mirrors, are placed near the transmit antenna. These RF mirrors controlled by information bits can create different channel fade realizations, which are used to form the channel modulation alphabet. The RF mirrors in MBM do not require complex RF hardware like mixers, filters, etc., and are therefore simple in terms of hardware. The transmit antenna is driven by one RF chain, resulting in a single RF chain implementation. Importantly, the rate increases linearly with the number of RF mirrors used in MBM. This is unlike in SM, which requires exponential increase in the number of transmit antennas elements to increase the rate. However, the RF mirrors add to the RF hardware complexity in MBM. An implementation of a compact MBM structure with 14 RF mirrors is reported in [14], where RF mirrors are implemented as periodic switched structures. MBM has been shown to achieve very good performance in the point-to-point setting [12]-[16]. Motivated by its hardware feasibility and superior performance in the point-to-point setting, we study MBM in the multiuser setting in this paper. In the rest of this section, we briefly present the basics of MBM.

The RF mirrors placed around a transmit antenna in MBM act as RF signal scatterers, which modify the propagation environment near the transmit antenna. Each RF mirror can be switched ON (reflects the incident RF signal from the transmit antennas) or OFF (transparent to the incident RF signal) digitally. The ON/OFF status of the mirrors is referred to as the 'mirror activation pattern' (MAP). If there are m_{rf} mirrors around a transmit antenna, then $2^{m_{rf}}$ MAPs are possible. Each of these MAPs corresponds to a unique near field geometry. In a rich scattering environment, even a small perturbation in the near field geometry results in an independent fading channel observed at the receiver in the far field. Thus, each MAP creates a corresponding independent fade realization. The transmitter can select one of the $2^{m_{rf}}$ MAPs using m_{rf} input information bits. The antenna element transmits a symbol from a conventional modulation alphabet (e.g., QAM) denoted by $\mathbb{A}$. Therefore, the achieved rate of MBM is given by $\eta_{\text{MBM}} = m_{rf} + \log_2 |\mathbb{A}|$ bpcu. The achieved rate increases by 1 bit with the addition of every single RF mirror. Thus, the MBM scheme not only assures a wireless channel with rich scattering and independent fade states, but also achieves high rates. A practical implementation of the MBM scheme with 14 RF mirrors and a dipole transmit antenna is reported in [14]. A scheme similar to MBM was reported earlier as 'aerial modulation' in [17], [18].

In addition to providing high rates, MBM has been shown to provide attractive performance advantages [12]-[16]. It has been shown that, MBM using n_r receive antennas can asymptotically (as $m_{rf} \to \infty$) achieve the capacity of n_r parallel AWGN channels [13]. This suggests that MBM can be an attractive candidate for use in massive MIMO systems, where the BS has a large number of receive antennas. Accordingly, we investigate MBM in the context of large-scale multiuser MIMO systems on the uplink.

First, considering small systems with few users and BS antennas, and maximum likelihood (ML) detection, we show that multiuser MBM (MU-MBM) can significantly outperform other multiuser MIMO systems that employ conventional modulation and spatial modulation. In order to scale MU-MBM to large systems with large number of users and BS antennas, we then proceed to investigate suitable techniques/algorithms meant for low-complexity signal detection and channel estimation at the BS receiver. We consider two approaches for this purpose. The first approach is based on the structured sparse signal recovery from compressive sensing theory and the second approach is based on message passing.

It has been recognized that compressive sensing (CS) will play a key role in the next generation 5G communications. This is mainly because of the rich set of tools provided by CS to exploit the sparsity, thereby enabling the recovery of high dimensional signals from a small number of linear observations (see [25], [26] and references therein for applications of CS in communications). In the present work, we recognize that multiuser MBM signal vectors are essentially sparse vectors, and therefore sparse signal recovery techniques are natural for MU-MBM signal detection. In this direction, we propose two CS based sparse vector reconstruction algorithms, namely, iterative sparse recovery (ISR) and inclusion exclusion subspace pursuit (IESP), for MU-MBM signal detection. ISR uses conventional CS based sparse recovery algorithms like orthogonal matching pursuit (OMP) [27], compressive sampling matching pursuit (CoSaMP) [28], and subspace pursuit (SP) [29] iteratively, with necessary modifications for the MU-MBM signal detection [32]. IESP is a structured sparse recovery algorithm, in which the conventional SP is enhanced to suit the structure of MU-MBM signal vectors to achieve low-complexity signal detection.

In the second approach based on message passing, we exploit 'channel hardening' – a phenomenon that occurs in large MIMO channels whereby the channel matrices in large-dimensional systems become well conditioned [33] – for the purpose of low-complexity detection and channel estimation in large-scale MU-MBM systems. In this approach, the multiuser MBM system is modeled as a bipartite graph and message passing is performed over it for efficient detection of the MU-MBM signal vectors [34]. Further, a low-complexity algorithm for combined channel estimation and detection in MU-MBM is proposed; this algorithm exploits channel hardening and enables MU-MBM signal detection at low computational complexity with estimated channel state information at the BS.

Both the proposed receiver approaches are shown to achieve very good performance in large-scale MU-MBM systems. For example, to achieve a certain target bit error performance, MU-MBM requires far fewer number of receive antennas at the BS compared to other traditional MIMO modulation techniques such as conventional modulation and spatial modulation. For

example, a bit error rate (BER) performance achieved using 620 receive antennas at the BS in a massive MIMO system using conventional modulation can be achieved using just 130 antennas with MU-MBM using the proposed structured sparse recovery algorithm. Also, spatial modulation used in the same massive MIMO settings requires more than 200 antennas to achieve the same BER performance. This suggests that MBM can be an attractive modulation scheme for use in the uplink of massive MIMO systems.

The rest of this paper is organized as follows. The MU-MBM system model and its BER performance with ML detection are presented in Sec. II. Sparsity-exploiting algorithms for MU-MBM signal detection and their performance in large-scale systems are presented in Sec. III. In Sec. IV, message passing based receivers for MU-MBM are presented. Conclusions are presented in Sec. V.

II. System model

Consider a multiuser MIMO system with K uplink users and a BS with n_r receive antennas (see Fig. 1). The users employ MBM for signal transmission. Each user has a single transmit antenna and m_{rf} RF mirrors placed near it. In a given channel use, each user selects one of the $2^{m_{rf}}$ possible mirror activation patterns (MAPs) using m_{rf} information bits. A MAP gives the ON/OFF status of each of the m_{rf} mirrors. Each of the $2^{m_{rf}}$ possible MAP is mapped to a bit sequence. An example of such a bit-to-MAP mapping is given in Table I for $m_{rf} = 2$.

Information bits	Mirror 1 status	Mirror 2 status
00	ON	ON
01	ON	OFF
10	OFF	ON
11	OFF	OFF

TABLE I: mapping between information bits and MAPs for $m_{rf} = 2$.

In addition to the bits conveyed through the choice of a MAP in a given channel use, a symbol from a modulation alphabet $\mathbb{A}$ (e.g., QAM, PSK) is also sent through the transmit antenna. This modulation symbol conveys an additional $\log_2 |\mathbb{A}|$ bits. Therefore, the spectral efficiency of a K-user MBM system is given by

$$\eta_{\text{MU-MBM}} = K(m_{rf} + \log_2 |\mathbb{A}|) \quad \text{bpcu}. \tag{1}$$

For example, a multiuser MBM system with $K = 4$, $m_{rf} = 2$, and 4-QAM has a total spectral efficiency of 16 bpcu. One of the key advantages of MBM is that the spectral efficiency per user increases linearly with the number of RF mirrors used at each user.

A. Single-user MBM channel alphabet

Let $M \triangleq 2^{m_{rf}}$, where M is the number of possible MAPs for m_{rf} RF mirrors. Let $\mathbf{h}_k^m$ denote the $n_r \times 1$ channel gain vector corresponding to the mth MAP of the kth user, where $\mathbf{h}_k^m = [h_{1,k}^m \; h_{2,k}^m \; \cdots \; h_{n_r,k}^m]^T$, $h_{i,k}^m$ is the channel gain between the kth user's transmit antenna and the ith receive antenna at the BS when the mth MAP is chosen at the transmitter,

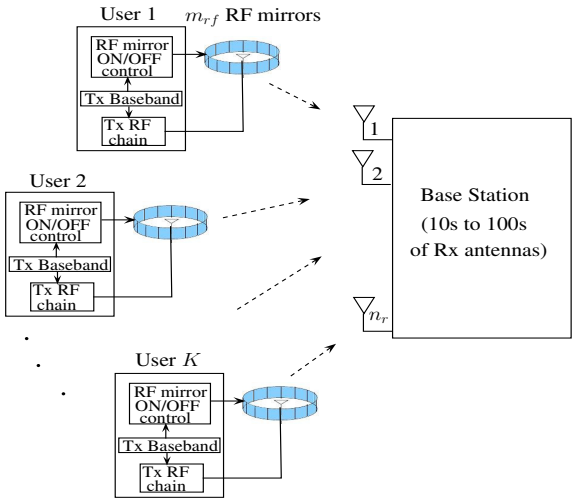


Fig. 1: Multiuser MBM in a massive MIMO system.

$i = 1, \cdots, n_r$, $k = 1, \cdots, K$, $m = 1, \cdots, M$, and the $h_{i,k}^m$s are assumed to be i.i.d. complex Gaussian random variables with mean zero and unit variance. The MBM channel alphabet for the kth user, denoted by $\mathbb{H}_k$, is defined as the collection of the channel gain vectors, i.e., $\mathbb{H}_k = \{\mathbf{h}_k^1, \mathbf{h}_k^2, \cdots, \mathbf{h}_k^M\}$. The MBM channel alphabet of each user is estimated at the BS receiver by transmitting known pilots before the data transmission phase. The MBM channel alphabet knowledge of each user is required at the BS receiver for detection. Whereas, the transmitters need not be aware of the channel alphabets.

B. Single-user MBM signal set

Define $\mathbb{A}_0 \triangleq \mathbb{A} \cup 0$. The single-user MBM signal set, denoted by $\mathbb{S}_{\text{SU-MBM}}$, is a set of $M \times 1$-sized MBM signal vectors given by

$$\mathbb{S}_{\text{SU-MBM}} = \left\{\mathbf{s}_{m,q} \in \mathbb{A}_0^M : m = 1, \cdots, M, \; q = 1, \cdots, |\mathbb{A}|\right\}$$
$$\text{s.t} \;\; \mathbf{s}_{m,q} = [0, \cdots, 0, \underbrace{s_q}_{m\text{th coordinate}} 0, \cdots, 0]^T, s_q \in \mathbb{A}, \tag{2}$$

where m is the index of the MAP. That is, an MBM signal vector $\mathbf{s}_{m,q}$ in (2) denotes that the complex symbol $s_q \in \mathbb{A}$ is transmitted on a channel whose gains are given by $\mathbf{h}^m$, where $\mathbf{h}^m$ is the $n_r \times 1$ channel gain vector corresponding to the mth MAP. Therefore, the $n_r \times 1$ received signal vector when the MBM signal vector $\mathbf{s}_{m,q}$ is transmitted is given by

$$\mathbf{y} = s_q \mathbf{h}^m + \mathbf{n}, \tag{3}$$

where $\mathbf{n} \in \mathbb{C}^{n_r}$ is the AWGN noise vector with $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_r})$. The size of the single-user MBM signal set is $|\mathbb{S}_{\text{SU-MBM}}| = M|\mathbb{A}|$. For example, if $m_{rf} = 2$ and $|\mathbb{A}| = 2$ (i.e., BPSK), then $|\mathbb{S}_{\text{SU-MBM}}| = 8$, and the corresponding MBM signal set is given by

$$\mathbb{S}_{\text{SU-MBM}} = \left\{ \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}-1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\-1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\-1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}, \begin{bmatrix}0\\0\\0\\-1\end{bmatrix} \right\}. \tag{4}$$

Note that all the M-length MU-MBM signal vectors have only one non-zero entry. The index of the non-zero entry determines the MAP used and the value of the non-zero entry gives the transmitted modulation symbol in the alphabet $\mathbb{A}$.

It has been shown in the point-to-point setting that, along with providing very high spectral efficiencies, MBM can also achieve good performance advantages. Further, it has been shown that, MBM with n_r receive antennas over a multipath fading channel asymptotically (as $m_{rf} \to \infty$) achieves the capacity of n_r parallel AWGN channels [13]. That is, in the limit as $m_{rf} \to \infty$, the capacity of MBM is

$$C = n_r \log(1+\gamma), \tag{5}$$

where γ is the signal-to-noise ratio (see Appendix A for proof). This suggests that MBM is an attractive choice for use in multiuser uplink systems having a large number of receive antennas at the BS.

C. Multiuser MBM signal set and received signal

The multiuser MBM signal set with K users is given by $\mathbb{S}_{\text{MU-MBM}} = \mathbb{S}^K_{\text{SU-MBM}}$. Let $\mathbf{x}_k \in \mathbb{S}_{\text{SU-MBM}}$ denote the transmit MBM signal vector from the kth user. Now, $\mathbf{x} = \begin{bmatrix}\mathbf{x}_1^T & \mathbf{x}_2^T & \cdots & \mathbf{x}_K^T\end{bmatrix}^T \in \mathbb{S}_{\text{MU-MBM}}$ is the vector comprising of the transmit signal vectors from all the K users. Let $\mathbf{H} \in \mathbb{C}^{n_r \times KM}$ denote the channel gain matrix given by $\mathbf{H} = [\mathbf{H}_1\ \mathbf{H}_2\ \cdots\ \mathbf{H}_K]$, where $\mathbf{H}_k = [\mathbf{h}_k^1\ \mathbf{h}_k^2\ \cdots\ \mathbf{h}_k^M] \in \mathbb{C}^{n_r \times M}$ is the MBM channel matrix of the kth user, and $\mathbf{h}_k^m$ is the channel gain vector of the kth user corresponding to the mth MAP. The entries of $\mathbf{H}$ are assumed to be distributed as i.i.d $\mathcal{CN}(0,1)$. The spatially correlated channel model and the effect of spatial correlation are presented in Sec. II-E. For this multiuser system, the $n_r \times 1$ received signal at the BS is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{6}$$

where $\mathbf{n}$ is the $n_r \times 1$ AWGN noise vector with $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$.

D. BER performance of MU-MBM with ML detection

In this subsection, we present the BER performance of MU-MBM under ML detection and compare it with that of other multiuser MIMO modulation techniques. The ML detection rule for the MU-MBM system model in (6) is given by

$$\widehat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{S}_{\text{MU-MBM}}}{\text{argmin}} \ \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2. \tag{7}$$

Figures 2a, 2b, and 2c show the BER performance of MU-MBM system with ML detection. In Fig. 2a, we consider a MU-MBM system with $K = 2$, $n_r = 8$, $m_{rf} = 3$, BPSK, and 4 bpcu per user. Let n_t and n_{rf} denote the number of transmit antennas and transmit RF chains, respectively, at each user. We compare the performance of this MU-MBM system with that of i) multiuser system with conventional modulation (MU-CM) with $n_t = 1$ and 16-QAM, and ii) multiuser system with spatial modulation (MU-SM) with $n_t = 2$ and 8-QAM. Note that in all the three systems, each user has only one transmit RF chain. All the schemes achieve a spectral efficiency of 4 bpcu per user. From Fig. 2a, we observe the following:

- The MU-MBM system achieves the best performance among all the three systems considered. For example, MU-MBM performs better by about 5 dB and 4 dB compared to MU-CM and MU-SM, respectively, at a BER of 10^{-5}.
- The better performance of MU-MBM can be attributed to more bits being conveyed through indexing RF mirrors. This allows MU-MBM to use lower-order modulation alphabets compared to other systems which may need higher-order alphabets to achieve the same bpcu.

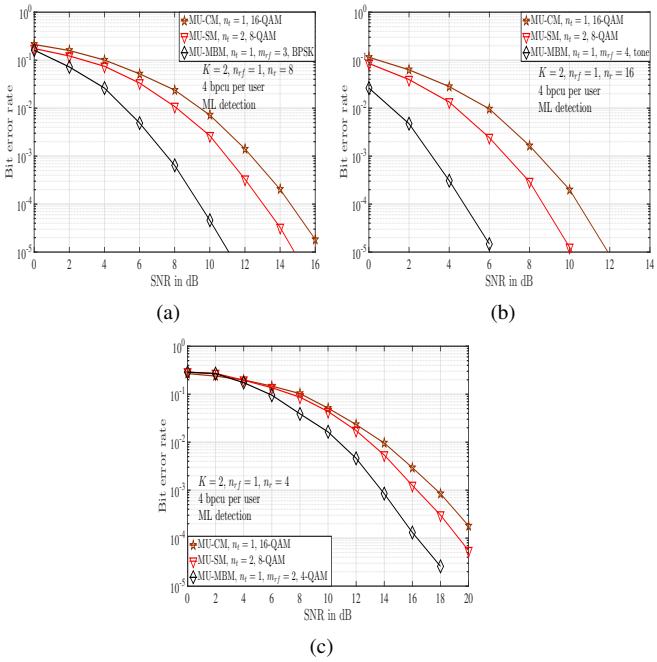


Fig. 2: BER performance of MU-MBM, MU-CM, and MU-SM with $K = 2$, 4 bpcu per user, and ML detection. (a) $n_r = 8$, (b) $n_r = 16$, and (c) $n_r = 4$.

Figure 2b shows the BER performance comparison of systems with $K = 2$ and $n_r = 16$. The systems compared are i) MU-MBM with $n_t = 1$, $m_{rf} = 4$, and tone, ii) MU-CM with $n_t = 1$, 16-QAM, and iii) MU-SM with $n_t = 2$, 8-QAM. All the systems achieve the rate of 4 bpcu per user. In Fig. 2c, the performance is compared among the systems with $K = 2$ and $n_r = 4$. The MU-CM and MU-SM systems have same configurations as in Fig. 2b, whereas MU-MBM uses $m_{rf} = 2$ and 4-QAM. All the three systems achieve the same rate of 4 bpcu per user. Performance trends similar to those observed in Fig. 2a are observed in Figs. 2b and 2c also.

E. Effect of spatial correlation

In generating the performance results in Fig. 2, the channel gain entries in $\mathbf{H}$ were assumed to be i.i.d. However, since the RF mirrors in a UE can be closely placed, there can be spatial correlation among the channels created by different MAPs. Further, there can be spatial correlation among the receive antennas. Also, because of the spatial separation between users, channels of different users towards the BS can be taken to be uncorrelated. That is, the entries of $\mathbf{H}_k$, for any given k, $k = 1, \cdots, K$, are assumed to be correlated, while the correlation among the entries of any pair of matrices $\mathbf{H}_k$ and $\mathbf{H}_l$, $k, l = 1, \cdots, K$ and $k \neq l$ is zero. We use the

Kronecker model for modeling channel correlation, i.e., the channel matrix with correlation is given by [15],[24]

$$\mathbf{H} = \mathbf{R}_{\text{Rx}}^{1/2} \tilde{\mathbf{H}} \mathbf{R}_{\text{Tx}}^{1/2}, \tag{8}$$

where $\mathbf{R}_{\text{Rx}}$ is the $n_r \times n_r$ receive correlation matrix, $\tilde{\mathbf{H}}$ is the $n_r \times KM$ matrix with entries distributed i.i.d $\mathcal{CN}(0,1)$, and $\mathbf{R}_{\text{Tx}}$ is the $KM \times KM$ transmit correlation matrix. The receive correlation matrix $\mathbf{R}_{\text{Rx}}$ is considered to follow exponentially decaying correlation model with its (i,j)th entry being $\rho_r^{|i-j|}$. With this, the receive correlation matrix $\mathbf{R}_{\text{Rx}}$ is given by

$$\mathbf{R}_{\text{Rx}} = \begin{bmatrix} 1 & \rho_r & \rho_r^2 & \cdots & \rho_r^{n_r-1} \\ \rho_r & 1 & \rho_r & \cdots & \rho_r^{n_r-2} \\ & & \ddots & & \\ \rho_r^{n_r-1} & \rho_r^{n_r-2} & & \cdots & 1 \end{bmatrix}. \tag{9}$$

The transmit correlation matrix $\mathbf{R}_{\text{Tx}}$ is given by

$$\mathbf{R}_{\text{Tx}} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}_K \end{bmatrix},$$

where $\mathbf{R}_k$ is the $M \times M$ RF mirror correlation matrix representing the correlation among the channels corresponding to different MAPs of kth user's RF mirrors. The correlation among MAPs of a given user is assumed to follow equicorrelation model with ρ_m being the correlation value among the different MAPs. With this model, the RF mirror correlation matrix $\mathbf{R}_k$, $k = 1, \cdots, K$ is given by

$$\mathbf{R}_k = \begin{bmatrix} 1 & \rho_m & \cdots & \rho_m \\ \rho_m & 1 & \cdots & \rho_m \\ & & \ddots & \\ \rho_m & \rho_m & \cdots & 1 \end{bmatrix}. \tag{10}$$

Figure 3 shows the effect of correlation on the BER performance of MU-MBM. The considered MU-MBM system has $K = 2$, $n_t = 1$, $m_{rf} = 2$, $n_r = 8$, 4-QAM, and 4 bpcu per user. We show the BER performance of this system for correlation values of $\rho_m = \rho_r = 0.4$, $\rho_m = \rho_r = 0.6$, and $\rho_m = \rho_r = 0.8$. It can be seen from Fig. 3 that, as expected, correlation degrades the BER performance, and that the degradation increases with increase in the amount of correlation. For example, at a BER of 10^{-3}, compared to the system with no correlation ($\rho_m = \rho_r = 0$), there is a degradation of about 1 dB at a correlation of $\rho_m = \rho_r = 0.4$. The degradation increases to about 3 dB and 7 dB at correlation values of $\rho_m = \rho_r = 0.6$ and $\rho_m = \rho_r = 0.8$, respectively.

Note that the ML detection complexity order is exponential in K. Hence, the complexity of ML detection is prohibitively high for systems with large K and n_r. However, in massive MIMO systems, K is in the order of tens and n_r is in the order of hundreds. Therefore, low complexity detection schemes that scale well for massive MU-MBM systems are essential. To address this need, we devise low-complexity detection algorithms for MU-MBM based on compressed sensing and message passing techniques in Sections III and IV, respectively.

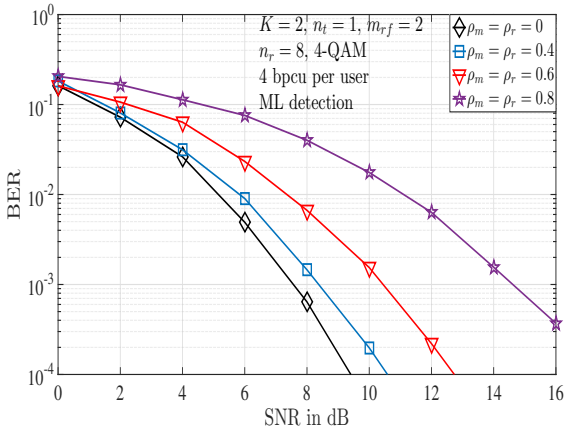


Fig. 3: Effect of correlation on the BER performance of MU-MBM system with $K = 2$, $n_t = 1$, $m_{rf} = 2$, $n_r = 8$, 4-QAM, and 4 bpcu per user, for correlation values of $\rho_m = \rho_r = 0, 0.4, 0.6$, and 0.8.

III. Sparsity-exploiting detection of multiuser MBM signals

In this section, we develop low-complexity signal detection algorithms based on compressed sensing techniques. These algorithms exploit the sparsity present in the MU-MBM signal vectors. An MBM signal vector has only one non-zero element out of M elements. This gives a sparsity factor of $1/M$ for MBM signals. For example, an MBM signal set with $m_{rf} = 4$ and $M = 2^{m_{rf}} = 16$ has a sparsity factor of $1/16$. Similarly, an MU-MBM signal vector for K users also has a sparsity factor of $K/KM = 1/M$. The MU-MBM signal vector comprises of K MBM sub-vectors, with each sub-vector having a single non-zero entry. This results in a structured sparsity referred to as the *'inclusion-exclusion sparsity'* [35]. Here, the inclusion of a non-zero entry in each sub-vector excludes all other entries of that sub-vector from being a non-zero value. We exploit this inherent structured sparsity in MU-MBM signals to devise algorithms that can lead to efficient signal detection at low computational complexities. We propose two different low-complexity MU-MBM signal detection algorithms based on the matching pursuit and subspace pursuit techniques known in compressive sensing.

First, we propose an algorithm that uses the greedy sparse recovery techniques known in compressed sensing such as orthogonal matching pursuit (OMP) [27], compressive sampling matched pursuit (CoSaMP) [28], and subspace pursuit(SP) [29]. We refer to this algorithm as the *iterative sparse recovery* (ISR). Next, we propose a detection algorithm that exploits the structured inclusion-exclusion sparsity in MU-MBM signal vectors for low-complexity detection. We refer to this algorithm as the *inclusion-exclusion subspace pursuit* (IESP) detector. IESP is an enhancement of SP algorithm, designed to suit the inclusion-exclusion structure of MU-MBM signal vectors. Another algorithm known as 'group subspace pursuit' has been reported in [31], which is also an extension of subspace pursuit to the distributed and group sparse signal structure of the uplink MU-SM in the multipath cyclic-prefix

single carrier channel when joint SM transmission is carried out. This work shows that, by extending the subspace pursuit algorithm to the distributed and group sparsity structure of the joint MU-SM transmission scheme, very good signal detection performance can be achieved.

A. The sparse signal recovery problem

The sparse signal recovery problem is finding an approximate solution to the following optimization problem

$$\arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}, \tag{11}$$

where $\mathbf{\Phi} \in \mathbb{C}^{m\times l}$ is referred to as the measurement matrix with $m < l$, $\mathbf{x} \in \mathbb{C}^l$ is the sparse signal vector, $\mathbf{y} \in \mathbb{C}^m$ is the noisy observation, and $\mathbf{n} \in \mathbb{C}^m$ is the noise. In conventional sparse recovery algorithms, the sparse vector $\mathbf{x}$ is not subjected to any further structural constraints. In compressed sensing based sparse vector reconstruction algorithms, the measurement matrix has to satisfy the restricted isometry property (RIP) for guaranteed reconstruction.

The MU-MBM signal detection problem at the BS can be modeled as a sparse recovery problem where the channel matrix $\mathbf{H} \in \mathbb{C}^{n_r\times KM}$ corresponds to the rectangular measurement matrix, the received signal vector $\mathbf{y} \in \mathbf{C}^{n_r}$ is the noisy observation, the noise is additive complex Gaussian $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$, and $\mathbf{x} \in \mathbb{S}_{\text{MU-MBM}}$ is the sparse vector that has to be reconstructed at the BS. In [36], it is shown that, for random matrices with i.i.d Gaussian entries, RIP is satisfied with very high probability. Thus, the channel matrix $\mathbf{H}$ satisfies RIP with very high probability.

B. MU-MBM detection as structured sparse recovery problem

The conventional sparse vector recovery problem does not pose any additional structural constraints, other than sparsity, on the signal vector. However, our signal of interest, i.e., the MU-MBM signal vector, is not only sparse, but also exhibits the inclusion-exclusion sparsity property. Further, the K non-zero entries in the sparse vector $\mathbf{x}$ can take values only from the alphabet $\mathbb{A}$. With these additional constraints, the signal detection problem in MU-MBM BS is to recover $\mathbf{x}$ from the observation vector

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \text{ s.t. } \|\mathbf{x}_k\|_0 = 1,\ s_k \in \mathbb{A}\ \forall k = 1, \cdots, K, \tag{12}$$

where $\mathbf{x}_k$ is the kth user's transmit MBM vector with $s_k \in \mathbb{A}$. In the following subsections, we present the proposed compressed sensing based detection algorithms to solve this detection problem.

C. Iterative sparse recovery (ISR)

The ISR algorithm consists of iteratively performing a sparse recovery algorithm (OMP or CoSaMP or SP) for the reconstruction of the MU-MBM signal vector till certain conditions are met. The listing of the proposed ISR based detection algorithm is given in **Algorithm 1**.

The SR in **Algorithm 1** denotes the sparse recovery algorithm, which can be any one of OMP, CoSaMP, and SP. The signal vector reconstructed by SR $(\cdot,\cdot,\cdot)$ is denoted by $\tilde{\mathbf{x}}$. Detecting the MU-MBM signal vector involves detecting the MBM signal vector transmitted by each user. An MBM signal vector from a user has exactly one non-zero entry out of M entries as observed in the example MBM signal set in (4). Hence, SR is expected to reconstruct a MU-MBM signal vector with the MBM sub-vector of each user having only one non-zero entry. But this constraint on the expected support set is not built in the conventional sparse recovery algorithms. In general, a sparse recovery algorithm can reconstruct K non-zeros at any of the KM locations of $\tilde{\mathbf{x}}$. This may result in reconstruction of an invalid signal vector. For example, $\text{SR}(\cdot,\cdot,\cdot)$ can reconstruct two non-zeros in the sub-vector of one user and no non-zeros in the sub-vector of another user. Though the former can be mapped to the nearest vector in $\mathbb{S}_{\text{SU-MBM}}$ based on the Euclidean distance, the latter can not be mapped. To overcome this issue, we define user activity pattern (UAP), denoted by $\mathbf{u}$, as a K-length vector with the kth entry $\mathbf{u}_k = 1$ if there is at least one non-zero entry in the kth user's sub-vector, and $\mathbf{u}_k = 0$ otherwise. A valid reconstructed signal vector is one which has all ones in $\mathbf{u}$, i.e., at least one non-zero is reconstructed in each user's sub-vector.

SR is initially performed with a sparsity estimate of K. If the reconstructed signal vector has a valid UAP, then each user's sub-vector is mapped to the nearest MBM vector and the algorithm terminates. If the reconstructed signal vector has an invalid UAP, then there can be more than one non-zeros in the sub-vectors of some users and all zeros in sub-vectors of some other users. In this case, the SR is performed multiple times with a range of sparsity estimates starting from K ($K+j$ in the algorithm listing, where j is the iteration variable) till the signal vector with the valid UAP is reconstructed (i.e., till the algorithm reconstructs at least one non-zero entry for each user's MBM signal vector). In the algorithm listing, $\mathbf{u}^j$ denotes the UAP at the jth iteration. It should be noted that, a signal vector with valid UAP will be reconstructed for $j < K(M-1)$, since at $j = K(M-1)$, the sparsity input $K + j = KM$, which is equal to the length of the MU-MBM signal vector.

Algorithm 1 Iterative sparse recovery

Inputs: $\mathbf{y}, \mathbf{H}, K$
Initialize: $j = 0$
repeat
  $\tilde{\mathbf{x}} = \text{SR}(\mathbf{y}, \mathbf{H}, K + j)$ ▷ Sparse Recovery algorithm
  $\mathbf{u}^j = \text{UAP}(\tilde{\mathbf{x}})$ ▷ Extract User Activity Pattern
  if $\|\mathbf{u}^j\|_0 = K$
    for $k = 1$ to K
      $\widehat{\mathbf{x}}_k = \underset{\mathbf{s}\in\mathbb{S}_{\text{SU-MBM}}}{\text{argmin}} \|\tilde{\mathbf{x}}_k - \mathbf{s}\|^2$ ▷ Nearest MBM signal mapping
    end for
    break;
  else $j = j + 1$
  end if
until $j < K(M-1)$
Output: The estimated MU-MBM signal vector

$$\widehat{\mathbf{x}} = [\widehat{\mathbf{x}}_1^T, \widehat{\mathbf{x}}_2^T, \cdots, \widehat{\mathbf{x}}_K^T]^T$$

However, it has been observed that, a signal vector with valid UAP is reconstructed for $j << K(M-1)$. On recovering an $\tilde{\mathbf{x}}$ with valid UAP, the reconstructed signal vector of each user is mapped to the nearest (in the Euclidean sense) MBM signal vector in $\mathbb{S}_{\text{SU-MBM}}$. This is shown in the Step 8 in the algorithm listing, where $\tilde{\mathbf{x}}_k$ denotes the recovered MBM signal vector of the kth user and $\widehat{\mathbf{x}}_k$ denotes the MBM signal vector to which $\tilde{\mathbf{x}}_k$ gets mapped to. Finally, the MU-MBM signal vector is obtained by concatenating the detected MBM signal vectors of all the users, i.e., , $\widehat{\mathbf{x}} = [\widehat{\mathbf{x}}_1^T, \widehat{\mathbf{x}}_2^T, \cdots, \widehat{\mathbf{x}}_K^T]^T$.

D. Inclusion-exclusion subspace pursuit (IESP)

The ISR detection algorithm requires multiple iterations of the SR algorithm to reconstruct the MU-MBM signal vector. The computational complexity of this compressed sensing based detection process can be further reduced if the inclusion-exclusion structure of the sparse MU-MBM signal is exploited. Using subspace pursuit (SP), we propose the inclusion-exclusion SP (IESP) detector, which exploits this structure to perform low complexity MU-MBM signal detection. Further, IESP has lesser computational complexity compared to ISR. The listing of the proposed IESP algorithm is presented in **Algorithm** 2.

Algorithm 2 Inclusion-exclusion subspace pursuit

1: Inputs: $\mathbf{y}, \mathbf{H}, K$
2: Initialize:
3: $\mathcal{S}^0 = \{l_1^0, l_2^0, \cdots, l_K^0\}$ s.t $l_k^0 = \text{argmax}_{j \in \mathcal{B}_k} \mathbf{h}_j^H \mathbf{y}$, $\forall k = 1, \cdots, K$
4: $\mathbf{a}^0 = \mathbf{H}_{\mathcal{S}^0}^{\dagger} \mathbf{y}$
5: $\mathbf{r}^0 = \mathbf{y} - \mathbf{H}_{\mathcal{S}^0} \mathbf{a}^0$
6: Iteration: In the ith iteration, do the following
7: $\tilde{\mathcal{S}}^i = \mathcal{S}^{i-1} \cup \{l_1', \cdots, l_K'\}$ s.t $l_k' = \text{argmax}_{j \in \mathcal{B}_k} \mathbf{h}_j^H \mathbf{r}^i, \forall k = 1, \cdots, K$
8: $\mathbf{z} = \mathbf{H}_{\tilde{\mathcal{S}}^i}^{\dagger} \mathbf{y}$
9: $\mathcal{S}^i = \{l_k^i = \text{argmax}_{l_k^{i-1}, l_k'} [z_{l_k^{i-1}}, z_{l_k'}], \quad \text{for } k = 1, \cdots, K\}$
10: $\mathbf{a}^i = \mathbf{H}_{\mathcal{S}^i}^{\dagger} \mathbf{y}$
11: $\mathbf{r}^i = \mathbf{y} - \mathbf{H}_{\mathcal{S}^i} \mathbf{a}^i$
12: If $\|\mathbf{r}^i\| > \|\mathbf{r}^{i-1}\|$, let $\mathcal{S}^i = \mathcal{S}^{i-1}$ and quit the iteration
13: $\mathcal{S} = \mathcal{S}^l$ and $\mathbf{a} = \mathbf{H}_{\mathcal{S}}^{\dagger} \mathbf{y}$
14: $s_k = \text{argmin}_{s \in \mathbb{A}} \|a_k - s\|^2$, $\forall k = 1, \cdots, K$
15: Output: The estimated MU-MBM signal vector $\widehat{\mathbf{x}}$ satisfying $\widehat{\mathbf{x}}_{\{1, \cdots, KM\} \setminus \mathcal{S}} = \mathbf{0}$ and $\widehat{\mathbf{x}}_{\mathcal{S}} = \mathbf{a}$.

Let $\mathcal{B}_k \triangleq \{(k-1)M+1, (k-1)M+2, \cdots, (k-1)M+M\}$. The set $\mathcal{B}_k$ denotes the set of all possible values of the index of the non-zero element in $\mathbf{x}_k$. Let $\mathcal{S}^i$ denote the set of the estimates of the non-zero indices of $\mathbf{x}$ at the ith iteration, l_k^i be the kth element of $\mathcal{S}^i$, and $|\mathcal{S}^i| = K$. In all iterations, we restrict the values of l_k^i to be from $\mathcal{B}_k$, i.e., $l_k^i \in \mathcal{B}_k$. This ensures the inclusion-exclusion sparsity structure in the reconstructed MU-MBM signal vector. We initialize $\mathcal{S}^0$, such that l_k^0 takes a value from the set $\mathcal{B}_k$ that maximizes $\mathbf{h}_{l_k^0}^H \mathbf{y}$, for $k = 1, 2, \cdots, K$. Let $\mathbf{r}^i \triangleq \mathbf{y} - \mathbf{H}_{\mathcal{S}^i} \mathbf{H}_{\mathcal{S}^i}^{\dagger} \mathbf{y}$; if the estimate $\mathcal{S}^i$ is not the same as the true support set of $\mathbf{x}$, then $\mathbf{r}^i$ is a non-zero vector. We iteratively try to minimize this residue $\|\mathbf{r}^i\|$.

In each iteration, a candidate support set $\tilde{\mathcal{S}}^i$ is constructed with a maximum of $2K$ indices. The set $\tilde{\mathcal{S}}^i$ is obtained by adding the indices l_k' to $\mathcal{S}^{i-1}$, $k = 1, \cdots, K$, where l_k' is an element from the set $\mathcal{B}_k$ that maximizes $\mathbf{h}_{l_k'}^H \mathbf{r}_i$. Let $\mathbf{z}$ be the projection of $\mathbf{y}$ on to $\mathbf{H}_{\tilde{\mathcal{S}}^i}$, i.e., $\mathbf{z} = \mathbf{H}_{\tilde{\mathcal{S}}^i} \mathbf{y}$. Now, we update the support set as $\mathcal{S}^i$ such that

$$l_k^i = \begin{cases} l_k^{i-1} & z_{l_k^{i-1}} > z_{l_k'} \\ l_k' & \text{else} \end{cases}.$$

The residue is computed at every iteration; the algorithm terminates when the current residue is higher than the previous residue. The final value of the set $\mathcal{S}^i$ is the detected support of the MU-MBM signal vector. The MU-MBM signal vector is reconstructed at the BS by obtaining $\hat{s}_k$ as

$$\hat{s}_k = \underset{s \in \mathbb{A}}{\text{argmin}} \|\mathbf{h}_{l_k^i}' \mathbf{y} - s\|^2,$$

where $\mathbf{h}_{l_k^i}'$ is the l_k^ith row of $\mathbf{H}^{\dagger}$. The decoding of the kth user's information bits from the detected MU-MBM signal vector consists of two parts: (1) the activated mirror index bits are decoded from the value of l_k^i, and (2) the modulation symbol bits are decoded from the detected $\hat{s}_k$ values.

Note 1: MU-SM has a similar signal structure as that of MU-MBM. A difference is that, in the case of MU-MBM, the length of the MBM sub-vectors can only be powers of 2. This is because, if m_{rf} is the number of RF mirrors used, then the length of the MBM vector is $2^{m_{rf}}$. However, there is no such length constraint for SM signal vectors. That is, the length of SM vectors is equal to the number of antennas per user, which need not be power of 2. Since IESP is independent of this small difference in the signal structure, it can also be used in the signal detection of MU-SM signal vectors.

Note 2: In [30], an algorithm similar to the IESP algorithm, called 'spatial modulation matching pursuit (SMMP)', is proposed in the context of generalized spatial modulation in multiple access. The following are some similarities and differences between SMMP and IESP.

- The SMMP detector is designed for the signal structure of generalized spatial modulation which allows more than one non-zero entry in the sub-vector corresponding to each user, depending on the number of active antennas per user. The IESP detector, on the other hand, considers a signal structure which can only have a single non-zero entry per user as per the structure of MU-MBM signal vector. However, SMMP can be specialized to the inclusion exclusion structure of the MU-MBM and hence can be used for MU-MBM signal detection.
- While SMMP is an extension of CoSaMP, IESP is based on SP. It has been shown in [29] that SP is superior to CoSaMP in terms of the reconstruction performance. Therefore, IESP which is an extension of SP is expected to achieve better performance compared to SMMP which is an extension of CoSaMP. We will see this in the simulation results in Sec. III-F.
- In any greedy sparse recovery algorithm, one of the key steps contributing to complexity is the least squares (LS) step used to obtain the values of the non-zero entries, after the support is recovered in a greedy manner. Since

SMMP is an extension of CoSaMP, the LS step in SMMP is computationally more expensive compared to IESP. While the IESP requires inversion of an $n_r \times K$ matrix in the LS step, the SMMP requires inversion of $n_r \times L$ $(L > K)$ matrix.

E. Computational complexity

The computational complexity of OMP, CoSaMP, and SP for an $m \times l$ matrix and a K-sparse vector is $O(mlK)$ [29]. From **Algorithm 1**, the computational complexity of the sparse vector reconstruction of the ISR algorithm is given by

$$O\Big(\underbrace{KM}_{\substack{\text{maximum number} \\ \text{of iterations}}} \times \underbrace{KM}_{\substack{\text{worst case} \\ \text{sparsity estimate}}} \times n_r \times KM \Big) = O(K^3 M^3 n_r).$$

The worst case complexity of ISR detector is cubic in the number of users and mirrors. It is observed from simulations that the worst case occurs only for low values of SNR. This is because, the SR algorithm is executed for several iterations to obtain a valid MU-MBM signal vector. However, the complexity of ISR detector is observed to be lesser than the worst case complexity at moderate to high values of SNR.

The order of the computational complexity of the IESP detector is same as that of SP algorithm. Therefore, the complexity of IESP detector is given by $O(K^2 n_r M)$, which is lesser than that of the ISR detector.

F. Performance of the proposed detectors

In this subsection, we present the BER performance of the proposed detectors for MU-MBM in a massive MIMO setting. Further, we compare this performance with that of other multiuser systems using conventional modulation (MU-CM) and spatial modulation (MU-SM) schemes.

Performance of the ISR detector: In Fig. 4, we show the BER performance of a MU-MBM system with $K = 16$ and $n_r = 128$. Each user uses $n_t = 1$, $m_{rf} = 6$, 4-QAM, spectral efficiency of 8 bpcu per user, and a sparsity factor of $1/64$. We compare the performance of the proposed ISR detector with that of MMSE detector. We consider ISR detector with three different SR algorithms: 1) OMP, 2) CoSaMP, and 3) SP. From Fig. 4, we observe that the ISR detector with OMP, CoSaMP, and SP algorithms achieve significantly better performance compared to MMSE. The ISR detector with SP algorithm outperforms all other detectors.

Performance comparison between ISR, IESP, SMMP, and MMSE detectors: In Fig. 5, we compare the performances of the proposed ISR (with SP) and IESP detectors with those of SMMP [30] and MMSE detectors. We consider MU-MBM with $K = 20$, $n_t = 1$, $n_r = 128$, $m_{rf} = 3$, 4-QAM, and 5 bpcu per user. We observe that the sparsity exploiting detectors (ISR, IESP, and SMMP detectors) perform better than the MMSE detector. Among the sparsity exploiting detectors, the IESP detector outperforms the ISR and SMMP detectors. The IESP detector achieves better performance than ISR detector since it exploits the inclusion-exclusion structure of the MU-MBM signal vector. The better performance of IESP compared to SMMP follows from the fact that SP has better reconstruction performance compared to CoSaMP [29].

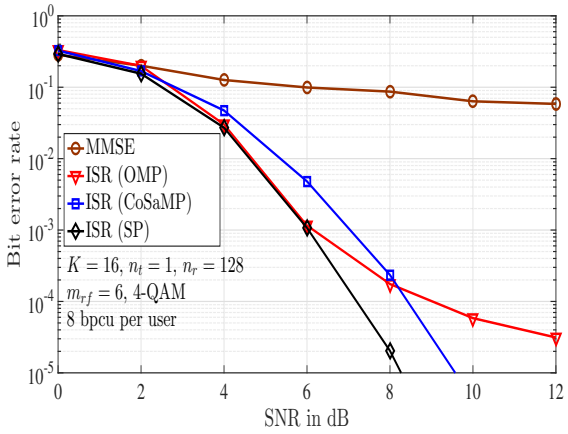


Fig. 4: BER performance of MU-MBM in a massive MIMO setting with $K = 16$, $n_r = 128$, $n_t = 1$, $m_{rf} = 6$, 4-QAM, 8 bpcu per user, using ISR detection. MMSE detection performance is also shown for comparison.

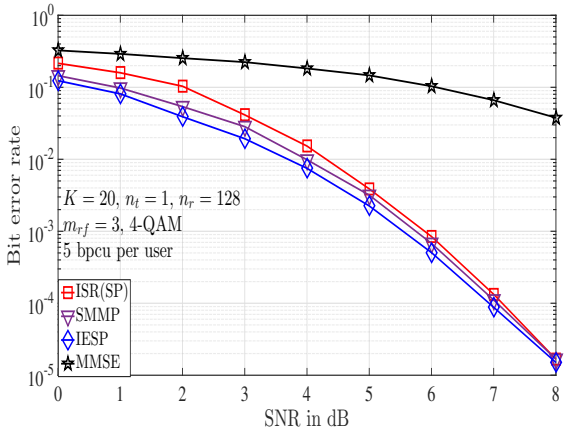


Fig. 5: BER Performance of MU-MBM with $K = 20$, $n_t = 1$, $n_r = 128$, $m_{rf} = 3$, 4-QAM, 5 bpcu per user using ISR (with SP), IESP, SMMP, and MMSE detectors.

Performance comparison with MU-CM and MU-SM: In Fig. 6, we compare the BER performances of i) MU-MBM with $n_t = 1$, $m_{rf} = 3$, and 4-QAM, ii) MU-CM with $n_t = 1$ and 32-QAM, and iii) MU-SM with $n_t = 4$ and 8-QAM. We consider all the systems in a massive MIMO setting with $K = 16$, $n_r = 128$, and a spectral efficiency of 5 bpcu per user. The IESP detector is used for detection of MU-MBM and MU-SM. Sphere decoder (ML detection) is used for MU-CM signal detection. The sparsity factors in MU-MBM and MU-SM are $1/8$ and $1/4$, respectively. It can be seen that the MU-MBM system clearly outperforms MU-CM and MU-SM systems. For example, at a BER of 10^{-5}, MU-MBM outperforms MU-CM and MU-SM by about 7 dB and 4 dB, respectively. This performance advantage of MU-MBM can be attributed to its signal distance properties [13]. MU-MBM also has the advantages of lower sparsity factor and using lower-order QAM size as additional bits are conveyed through indexing mirrors.

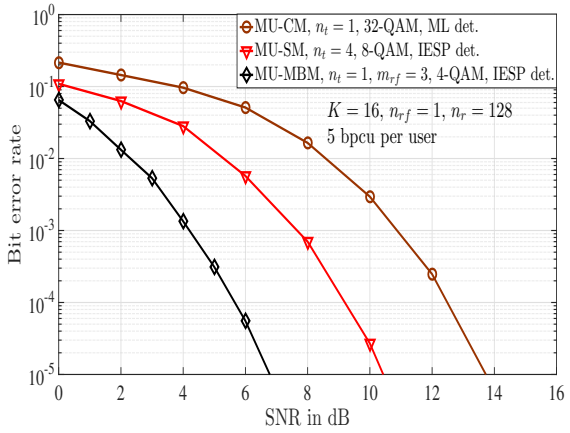


Fig. 6: BER performance of MU-MBM, MU-CM, and MU-SM in a massive MIMO setting with $K = 16$, $n_r = 128$, and 5 bpcu per user.

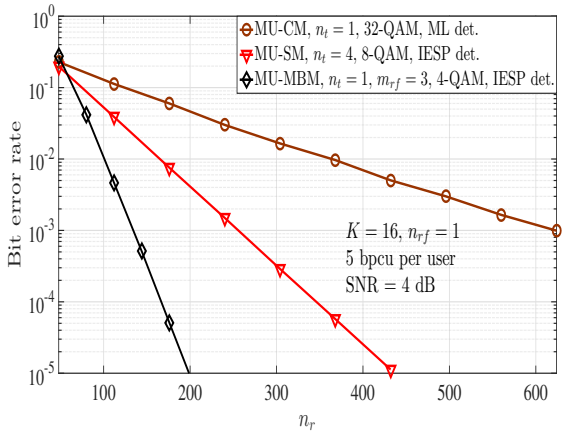


Fig. 7: BER performance of MU-MBM, MU-CM, and MU-SM as a function of n_r in a massive MIMO setting with $K = 16$, 5 bpcu per user, and SNR = 4 dB.

Effect of the number of receive antennas: Here, we analyze the performance of MU-MBM, MU-CM, and MU-SM systems as the number of BS receive antennas is increased (Fig. 7). We consider systems with $K = 16$, 5 bpcu per user, SNR = 4 dB, and n_r is varied from 48 to 624. The parameters of the systems are the same as described before. It is observed that the MU-MBM system requires only 130 receive antennas to achieve a BER of 3×10^{-3}; whereas, the MU-CM system requires 620 receive antennas to achieve the same BER performance. To achieve a BER of 10^{-5}, the MU-SM also requires 230 receive antennas more than the number of antennas in MU-MBM to achieve the same BER. This advantage of MU-MBM system can be mainly attributed to its better signal distance properties, particularly when n_r is large [13]. Thus, we see that the multiuser MBM is a promising candidate for the uplink in massive MIMO systems.

IV. Message passing based MU-MBM signal detection

It is known that message passing based detection algorithms can achieve very good performance in large-dimensional MIMO systems [20]-[22]. In this section, we present message passing based MU-MBM signal detectors. In the first detection algorithm, we model the MU-MBM system using a bipartite probabilistic graph and perform inference over this graph using belief propagation to detect the MBM signals. Next, we develop a receiver that exploits the channel hardening phenomenon that occurs in large MIMO channels.

A. BP based MBM signal detector (BP-MSD)

The MU-MBM system model described in (6) can be represented using a bipartite graph as illustrated in Fig. 8. This bipartite graph consists of K variable nodes, each corresponding to a user's transmit vector $\mathbf{x}_j$, and n_r observation nodes, each corresponding to a received signal value y_i. We develop a BP based detection algorithm that estimates $\mathbf{x}$ given the observation $\mathbf{y}$ and the channel matrix $\mathbf{H}$. Here, we assume that the BS has perfect knowledge of $\mathbf{H}$. From (6), the received signal y_i can be written as

$$y_i = \mathbf{h}_{i,[j]}\mathbf{x}_j + \underbrace{\sum_{l=1, l\neq j}^{K} \mathbf{h}_{i,[l]}\mathbf{x}_l + n_i}_{\triangleq\, q_{i,j}}, \tag{13}$$

where $\mathbf{h}_{i,[l]}$ is a row vector of length M, given by $\begin{bmatrix} H_{i,(l-1)M+1} & H_{i,(l-1)M+2} & \cdots & H_{i,lM} \end{bmatrix}$, $H_{i,j}$ is the (i,j)th entry of the matrix $\mathbf{H}$, and $q_{i,j}$ is the total interference in the ith received signal. In large-scale multiuser MBM systems, the value of KM is large. Hence, we employ Gaussian approximation of interference. That is, we approximate the interference term $q_{i,j}$ to be Gaussian with mean $\mu_{i,j}$ and variance $\sigma^2_{i,j}$, where

$$\mu_{i,j} = \mathbb{E}\Big[\sum_{l=1,l\neq j}^{K} \mathbf{h}_{i,[l]}\mathbf{x}_l + n_i\Big] = \sum_{l=1,l\neq j}^{K} \sum_{\mathbf{s}\in\mathbb{S}_{\text{SU-MBM}}} p_{li}(\mathbf{s})\mathbf{h}_{i,[l]}\mathbf{s}, \tag{14}$$

$$\begin{aligned}\sigma^2_{i,j} &= \text{Var}\Big(\sum_{\substack{l=1,\\ l\neq j}}^{K} \mathbf{h}_{i,[l]}\mathbf{x}_l + n_i\Big) \\ &= \sum_{\substack{l=1,\\ l\neq j}}^{K}\Big(\sum_{\mathbf{s}\in\mathbb{S}_{\text{SU-MBM}}} p_{li}(\mathbf{s})\mathbf{h}_{i,[l]}\mathbf{s}\mathbf{s}^H\mathbf{h}^H_{i,[l]} \\ &\quad - \Big|\sum_{\mathbf{s}\in\mathbb{S}_{\text{SU-MBM}}} p_{li}(\mathbf{s})\mathbf{h}_{i,[l]}\mathbf{s}\Big|^2\Big) + \sigma^2,\end{aligned} \tag{15}$$

where $p_{ji}(\mathbf{s})$ is the message passed by the jth variable node to the ith observation node. The a posteriori probability (APP) or 'belief' $p_{ji}(\mathbf{s})$ is defined as $p(\mathbf{x}_j = \mathbf{s}|\mathbf{y}_{\setminus i})$, where $\mathbf{y}_{\setminus i}$ denotes the vector of all elements in $\mathbf{y}$ except y_i. The message $p_{ji}(\mathbf{s})$ can be evaluated as

$$p_{ji}(\mathbf{s}) \propto \prod_{m=1,\ m\neq i}^{n_r} \exp\left(\frac{-\left|y_m - \mu_{m,j} - \mathbf{h}_{m,[j]}\mathbf{s}\right|^2}{\sigma^2_{m,j}}\right). \tag{16}$$

The belief propagation schedule is given below.

1) Initialize $p_{ji}(\mathbf{s}) = \frac{1}{|\mathbb{S}_{\text{SU-MBM}}|}$, $\forall j, i, \mathbf{s}$.
2) Compute μ_{ij} and $\sigma^2_{i,j}$, $\forall i, j$.
3) Compute p_{ji}, $\forall j, i$.

The belief values computed through (16) are damped with a damping factor $\beta \in (0, 1]$. It is known that damping of beliefs

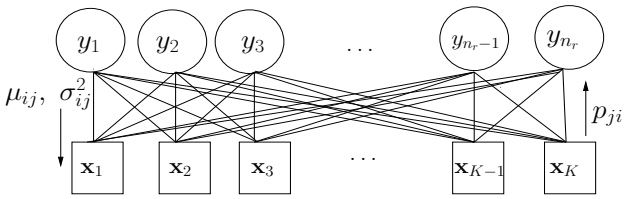


Fig. 8: Graphical model for BP-MSD.

improves the convergence of the message passing algorithm [37]. The steps 2 and 3 are repeated till the APP values converge or a maximum number of iterations is reached. When the algorithm is terminated, the probabilities $p(\mathbf{x}_j = \mathbf{s})$, $j = 1, 2, \cdots, K$, are computed as

$$p_j(\mathbf{s}) \propto \prod_{i=1}^{n_r} \exp\left(\frac{-\left|y_i - \mu_{i,j} - \mathbf{h}_{i,[j]}\mathbf{s}\right|^2}{\sigma^2_{i,j}} \right), \tag{17}$$

Finally, the estimates $\widehat{\mathbf{x}}_j$s are obtained by choosing the MBM signal vector $\mathbf{s} \in \mathbb{S}$ that has the largest APP. That is,

$$\widehat{\mathbf{x}}_j = \underset{\mathbf{s} \in \mathbb{S}_{\text{SU-MBM}}}{\text{argmax}}\ p_j(\mathbf{s}). \tag{18}$$

The $\widehat{\mathbf{x}}_j$s obtained are demapped to get the MAP, which are then demapped to obtain the bits corresponding to the MAP of each user. The non-zero entry in $\widehat{\mathbf{x}}_j$ is demapped to get the conventional modulation symbol bits transmitted by each of the K users. The listing of the steps in BP-MSD is given in **Algorithm 3**.

Algorithm 3 Listing of the BP-MSD algorithm

Input : $\mathbf{y}$, $\mathbf{H}$, σ^2; β: damping factor
Initialize: $p_{ji}^{(0)}(\mathbf{s}) \leftarrow \frac{1}{|\mathbb{S}|},\ \forall i, j, \mathbf{s}$.
for $t = 1 \rightarrow number\_of\_iterations$ **do**
  for $i = 1 \rightarrow N$ **do**
    for $j = 1 \rightarrow K$ **do**
      $\tilde{\mu}_{ij} \leftarrow \sum_{\mathbf{s} \in \mathbb{S}} p_{ji}^{(t-1)}(\mathbf{s}) \sum_{l \in \mathcal{I}(\mathbf{s})} s_l H_{i,(j-1)n_t+l}$
      $\tilde{\sigma}^2_{ij} \leftarrow \sum_{\mathbf{s} \in \mathbb{S}} p_{ji}^{(t-1)}(\mathbf{s}) \mathbf{h}_{i,[j]} \mathbf{s}\mathbf{s}^H \mathbf{h}^H_{i,[j]} - |\tilde{\mu}_{ij}|^2$
    end for
    $\mu_i \leftarrow \sum_{j=1}^{K} \tilde{\mu}_{ij}$
    $\sigma^2_{ij} \leftarrow \sum_{j=1}^{K} \tilde{\sigma}^2_{ij} + \sigma^2$
    for $j = 1 \rightarrow K$ **do**
      $\mu_{ij} \leftarrow \mu_i - \tilde{\mu}_{ij}$
      $\sigma^2_{ij} \leftarrow \sigma^2_i - \tilde{\sigma}^2_{ij}$
    end for
  end for
  for $j = 1 \rightarrow K$ **do**
    for $\mathbf{s} \in \mathbb{S}$ **do**
      $\ln(p_j^{(t)}(\mathbf{s}) \leftarrow \sum_{i=1}^{N} \frac{-|y_i - \mu_{ij} - \mathbf{h}_{i,[j]}\mathbf{s}|^2}{\sigma^2_{ij}}$
    end for
    for $i = 1 \rightarrow N$ **do**
      for all $\mathbf{s} \in \mathbb{S}$ **do**
        $\tilde{p}_{ji}^{(t)}(\mathbf{s}) \leftarrow \ln(p_{(t)}(\mathbf{s})) + \frac{|y_i - \mu_{ij} - \mathbf{h}_{i,[j]}\mathbf{s}|^2}{\sigma^2_{ij}}$
        $p_{ji}^{(t)} = \frac{1-\beta}{C_{ji}} \exp(\tilde{p}_{ji}^{(t)}(\mathbf{s})) + \beta p_{ji}^{(t-1)}(\mathbf{s})$
        (C_{ji} is a normalizing constant)
      end for
    end for
  end for
end for
Output : $p_j(\mathbf{s})$ and $\widehat{\mathbf{x}}_j = \underset{\mathbf{s} \in \mathbb{S}}{\text{argmax}}\, p_j(\mathbf{s}),\ \forall j$

B. Channel hardening-exploiting message passing for MU-MBM signal detection (CHEMP-MSD)

Large MIMO channels are known to exhibit the channel hardening phenomenon [20], this phenomenon can be explained as follows. In an $m \times l$ channel matrix, when m and l are increased, with their ratios fixed, the distribution of the singular values of the matrix becomes less sensitive to the actual distribution of the entries of the matrix (as long as the entries are i.i.d.) [1], [20], [33]. This is referred to as the channel hardening phenomenon. Because of the channel hardening effect, very tall or very wide channel matrices become very well conditioned. As a result, the off-diagonal terms of the $\mathbf{H}^H\mathbf{H}$ matrix become increasingly weaker compared to the diagonal terms as the size of the channel matrix $\mathbf{H}$ increases. We exploit this phenomenon to develop a message passing algorithm that can efficiently detect large-scale MU-MBM signals. We refer to this algorithm as the channel hardening-exploiting message passing for MU-MBM signal detection (CHEMP-MSD).

In order to exploit the channel hardening effect, the detection algorithm works on the matched filtered received signal vector, which is obtained as

$$\mathbf{H}^H\mathbf{y} = \mathbf{H}^H(\mathbf{H}\mathbf{x} + \mathbf{n}). \tag{19}$$

In the above equation, since the channel matrix $\mathbf{H}$ is a large matrix with i.i.d. complex Gaussian entries, it exhibits channel hardening phenomenon. Thus, $\mathbf{H}^H\mathbf{H}$ has strong diagonal components and weak off-diagonal elements. After normalizing by n_r, we can rewrite (19) as

$$\mathbf{z} = \mathbf{G}\mathbf{x} + \mathbf{w}, \tag{20}$$

$$\mathbf{z} \triangleq \frac{\mathbf{H}^H\mathbf{y}}{n_r}, \quad \mathbf{G} \triangleq \frac{\mathbf{H}^H\mathbf{H}}{n_r}, \quad \mathbf{w} \triangleq \frac{\mathbf{H}^H\mathbf{n}}{n_r}. \tag{21}$$

As in (6), $\mathbf{z}$ in (20) constitutes K sub-vectors, each of length M, i.e., $\mathbf{z} = [\mathbf{z}_1^T \mathbf{z}_2^T \cdots \mathbf{z}_i^T \cdots \mathbf{z}_K^T]^T$. Similarly, $\mathbf{w} = [\mathbf{w}_1^T \mathbf{w}_2^T \cdots \mathbf{w}_i^T \cdots \mathbf{w}_K^T]^T$, where $w_j = \sum_{l=1}^{n_r} \frac{H_{lj}^* n_l}{n_r}$ is the jth element of $\mathbf{w}$ and H_{ji} is the (j,i)th element of $\mathbf{H}$. Using central limit theorem, for large values of n_r (tens to hundreds), w_j can be approximated to be Gaussian with zero mean and variance $\sigma_w^2 \triangleq \frac{\sigma^2}{n_r}$. We also employ Gaussian approximation of interference in the computation of the messages in the message passing algorithm. From (20), $\mathbf{z}_i$ can be expressed as

$$\mathbf{z}_i = \mathbf{G}_{ii}\mathbf{x}_i + \underbrace{\sum_{j=1, j\neq i}^{K} \mathbf{G}_{ij}\mathbf{x}_j + \mathbf{w}_i}_{\triangleq\ \mathbf{g}_i}, \tag{22}$$

where $\mathbf{G}_{ij}$ is a $M \times M$ sub-matrix of $\mathbf{G}$ formed by taking the elements in rows $(i-1)M+1$ to iM and columns $(j-1)M+1$

to jM. The matrix $\mathbf{G}$ can thus be written in terms of the sub-matrices as

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} & \cdots & \mathbf{G}_{1K} \\ \mathbf{G}_{21} & \mathbf{G}_{22} & \cdots & \mathbf{G}_{2K} \\ \vdots & & \ddots & \vdots \\ \mathbf{G}_{K1} & \mathbf{G}_{K2} & \cdots & \mathbf{G}_{KK} \end{bmatrix}. \tag{23}$$

The interference-plus-noise term $\mathbf{g}_i$, for the ith user, is formed by the off-diagonal elements of $\frac{\mathbf{H}^H\mathbf{H}}{n_r}$ (i.e., $\mathbf{G}_{ij}$, $i \neq j$). Due to the channel hardening effect, the matrix $\mathbf{G}$ has strong diagonal elements compared to the off-diagonal elements for large n_r and K. We approximate $\mathbf{g}_i$ to be a multivariate Gaussian random vector with mean $\boldsymbol{\mu}_i$ and variance $\boldsymbol{\Sigma}_i$, which can be obtained as

$$\boldsymbol{\mu}_i \;=\; \mathbb{E}(\mathbf{g}_i) = \sum_{j=1, j\neq i}^{K} \mathbf{G}_{ij}\mathbb{E}(\mathbf{x}_j), \tag{24}$$

$$\boldsymbol{\Sigma}_i \;=\; \mathrm{Var}(\mathbf{g}_i) = \sum_{j=1, j\neq i}^{K} \mathbf{G}_{ij}\mathrm{Var}(\mathbf{x}_j)\mathbf{G}_{ij}^H + \sigma_w^2\mathbf{I}_M. \tag{25}$$

Let $\mathbf{p}_i$ denote the $|\mathbb{A}|M \times 1$ vector of probability values corresponding to the MBM signal vector $\mathbf{x}_i$. The entries of $\mathbf{p}_i$ are given by

$$p_i(\mathbf{s}) = \Pr\big(\mathbf{x}_i = \mathbf{s}\big),\ \mathbf{s} \in \mathbb{S}_{n_t,\mathbb{A}}. \tag{26}$$

Now, we have

$$\mathbb{E}(\mathbf{x}_j) \;=\; \sum_{\forall \mathbf{s},\ \mathbf{s}\in\mathbb{S}_{n_t,\mathbb{A}}} \mathbf{s}p_j(\mathbf{s}) \tag{27}$$

$$\mathrm{Var}(\mathbf{x}_j) \;=\; \sum_{\forall \mathbf{s},\ \mathbf{s}\in\mathbb{S}_{n_t,\mathbb{A}}} \mathbf{s}\mathbf{s}^H p_j(\mathbf{s}) - \mathbb{E}(\mathbf{x}_j)\mathbb{E}(\mathbf{x}_j)^H. \tag{28}$$

We approximate $\mathbf{p}_i$s with the corresponding APPs, i.e.,

$$p_i(\mathbf{s}) \leftarrow \Pr(\mathbf{x}_i = \mathbf{s}|\mathbf{z}_i, \mathbf{G}), \tag{29}$$

where

$$\Pr(\mathbf{x}_i = \mathbf{s}|\mathbf{z}_i, \mathbf{G}) \;\propto\; e^{-\frac{1}{2}(\mathbf{z}_i - \mathbf{G}_{ii}\mathbf{s} - \boldsymbol{\mu}_i)^H \boldsymbol{\Sigma}_i^{-1}(\mathbf{z}_i - \mathbf{G}_{ii}\mathbf{s} - \boldsymbol{\mu}_i)}. \tag{30}$$

Message passing: In CHEMP-MSD, the MU-MBM system model can be represented by a fully-connected graph with K nodes, where the ith node computes an approximate APP value corresponding to the ith user's transmit vector $\mathbf{x}_i$. Each node computes the APP using the incoming messages, the matrix $\mathbf{G}$ and the vector $\mathbf{z}_i$. The message passing schedule is as follows.

1) Initialize the probability vectors $\mathbf{p}_i$s with equiprobable values.
2) Each node computes the probabilities $\mathbf{p}_i$ as per (29) using (24) and (25).

The steps 1 and 2 are repeated till the probability values converge or till a maximum number of iterations is reached. As before, the messages are damped with a damping factor β to improve the convergence rate [37]. At the end of the tth iteration, the messages are damped with a damping factor $\beta \in [0, 1)$. Thus, if $\tilde{\mathbf{p}}_i^t$ is the computed probability vector at the tth iteration, the message at the end of tth iteration is given by

$$\mathbf{p}_i^t = (1-\beta)\tilde{\mathbf{p}}_i^t + \beta\mathbf{p}_i^{t-1}. \tag{31}$$

Algorithm 4 Listing of the CHEMP-MSD algorithm

Input : $\mathbf{z}$, $\mathbf{G}$, σ^2; β: damping factor
Initialize: $p_i^{(0)}(\mathbf{s}) \leftarrow \frac{1}{|\mathbb{S}|},\ i = 1, \cdots, K,\ \forall \mathbf{s} \in \mathbb{S}$.
for $t = 1 \rightarrow number\_of\_iterations$ **do**
  for $i = 1 \rightarrow K$ **do**
    $\mathbf{E}(\mathbf{x}_i) \leftarrow \sum_{\mathbf{s}\in\mathbb{S}} \mathbf{s}p_i^{(t-1)}(\mathbf{s})$
    $\mathrm{Cov}(\mathbf{x}_i) \leftarrow \sum_{\mathbf{s}\in\mathbb{S}} \mathbf{s}\mathbf{s}^H p_i^{(t-1)}(\mathbf{s}) - \mathbb{E}(\mathbf{x}_i)\mathbb{E}(\mathbf{x}_i)^H$
  end for
  for $i = 1 \rightarrow K$ **do**
    $\boldsymbol{\mu}_i \leftarrow \sum_{j=1, j\neq i}^{K} \mathbf{G}_{ij}\mathbb{E}(\mathbf{x}_j)$
    $\boldsymbol{\Sigma}_i \leftarrow \sum_{j=1, j\neq i}^{K} \mathbf{G}_{ij}\mathrm{Cov}(\mathbf{x}_j)\mathbf{G}_{ij}^H + \sigma_w^2\mathbf{I}_{N_m}$
    for all $\mathbf{s} \in \mathbb{S}$ **do**
      $\tilde{p}_i^t(\mathbf{s}) \leftarrow \frac{1}{C_i} \exp\big(-(\mathbf{z}_i - \mathbf{G}_{ii}\mathbf{s} - \boldsymbol{\mu}_i)^H \boldsymbol{\Sigma}_i^{-1}(\mathbf{z}_i - \mathbf{G}_{ii}\mathbf{s} - \boldsymbol{\mu}_i)\big)$
      (C_i is a normalizing constant)
      $p_i^t(\mathbf{s}) = (1-\beta)\tilde{p}_i^t(\mathbf{s}) + \beta p_i^{(t-1)}(\mathbf{s})$
    end for
  end for
end for
Output : $\mathbf{p}_i(\mathbf{s})$

After the iterative algorithm is terminated, an estimate of the conventional modulation symbol transmitted by the kth user is obtained as

$$\widehat{s}_k = \underset{s\in\mathbb{A}}{\mathrm{argmax}} \sum_{\forall \mathbf{s},\ \mathbf{s}\in\mathbb{S}_{n_t,\mathbb{A}} : \mathcal{X}(\mathbf{s})=s} p_k(\mathbf{s}), \tag{32}$$

where $\mathcal{X}(\mathbf{s})$ gives the non-zero element in $\mathbf{s}$. An estimate of the MAP chosen for transmission by the kth user is obtained as

$$\widehat{q}_k = \underset{q\in\{1,\cdots,N_m\}}{\mathrm{argmax}} \sum_{\forall \mathbf{s},\ \mathbf{s}\in\mathbb{S}_{n_t,\mathbb{A}} : \mathcal{I}(\mathbf{s})=q} p_k(\mathbf{s}), \tag{33}$$

where $\mathcal{I}(\mathbf{s})$ gives the index of the non-zero element in $\mathbf{s}$. The estimated MAP and the modulation symbols obtained are then demapped to obtain the kth user's bits. The listing of the steps involved in CHEMP-MSD is given in **Algorithm 4**.

C. MBM signal detection with estimated CSI

In this subsection, we consider a practical scenario where perfect channel knowledge is not available at the BS. Hence, the channel matrix has to be estimated at the BS to detect the MU-MBM signals. We refer to the system with a combined channel estimator and detector as a receiver. Here, we extend the message passing based detectors described in the previous section to develop large-scale MU-MBM signal receivers.

Here, the transmission happens in frames. Let the length of the transmission frame be F_L. We assume that the coherence time of the channel to be equal to the duration of the frame (hence, the channel becomes invariant over a frame duration). In each frame, the users transmit a pilot part and a data part. Pilot symbols are transmitted during the pilot part. The pilot part has KM channel uses, and the data part consists of $F_L - KM$ channel uses. Let $\mathbf{X}_\mathrm{p} = A\mathbf{I}_{KM}$ denote the pilot matrix. In the ith channel use, $1 \leq i \leq KM$, the $\lceil \frac{i}{M} \rceil$th user transmits a pilot symbol using a MAP, whose index is given by $\big((i-1) \mod M\big) + 1$, with amplitude A. Only one user transmits at a

time and all other users remain silent. The received signal at the BS during the pilot phase is given by

$$\mathbf{Y}_{\text{p}} \;=\; \mathbf{H}\mathbf{X}_{\text{p}} + \mathbf{N}_{\text{p}} \;=\; A\mathbf{H} + \mathbf{N}_{\text{p}}, \tag{34}$$

where $A = \sqrt{KE_s}$, E_s is the average symbol energy, and $\mathbf{N}_{\text{p}}$ denotes the noise matrix.

1) BP receiver: We extend the proposed BP-MSD algorithm to formulate the belief propagation based receiver (BP receiver). Using an MMSE channel estimator, we estimate the channel matrix $\mathbf{H}$. The estimated channel matrix is then used instead of $\mathbf{H}$ in the BP-MSD algorithm as described in the Section IV-A.

2) CHEMP receiver: This CHEMP receiver employs the detection algorithm described in Section IV-B for the detection of MBM signals. However, the CHEMP receiver, does not use the conventional method of estimating the channel matrix $\mathbf{H}$ directly (as in MMSE channel estimation). Instead, the CHEMP receiver directly obtains an estimate of $\mathbf{H}^H\mathbf{H}$ (which is defined as $\mathbf{G}$ in Section IV-B). This is because, the equivalent system model obtained in (20) through the matched filtering operation can work with a direct estimate of $\mathbf{G}$ without the knowledge of $\mathbf{H}$. We show that this approach of direct estimation of $\mathbf{G}$ performs better compared to the conventional method of explicitly estimating $\mathbf{H}$ and detecting $\mathbf{x}$. In the CHEMP receiver, the matrix $\mathbf{G}$ is estimated as

$$\widehat{\mathbf{G}} \;=\; \frac{\mathbf{Y}_{\text{p}}^T\mathbf{Y}_{\text{p}}}{n_r A^2} - \frac{\sigma_w^2}{A^2}\mathbf{I}_{KM}. \tag{35}$$

An estimate of the vector $\mathbf{z}$ is obtained as

$$\widehat{\mathbf{z}} \;=\; \frac{\mathbf{Y}_{\text{p}}^T\mathbf{y}}{n_r A}, \tag{36}$$

where $\mathbf{y}$ is the received signal vector in each channel use of the data part. These estimates $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{z}}$ obtained from (35) and (36) are used in place of $\mathbf{G}$ and $\mathbf{z}$ in the CHEMP detection algorithm.

D. Computational complexity

In BP-MSD, the computation of the messages from the observation nodes to the variable nodes has an order of $O(n_r M^2 K^2 |\mathbb{S}_{\text{SU-MBM}}|)$. The messages from the variable nodes to the observation nodes require a computational complexity order of $O(n_r^2 MK|\mathbb{S}_{\text{SU-MBM}}|)$.

In CHEMP-MSD, the complexity order for the computation of the APP is given by $O(M^3K(|\mathbb{S}_{\text{SU-MBM}}| + K))$. The computation of $\mathbf{G}$ has a complexity order of $O(n_r K^2 M^2)$. For $n_r > K$ and $K > M$, the overall complexity of the algorithm is dominated by the computation of $\mathbf{G}$.

The CHEMP receiver has no additional computational complexity for channel estimation. That is, the CHEMP receiver has the same computational complexity as the CHEMP-MSD algorithm despite performing channel estimation. In contrast, other receiver algorithms require additional computations to explicitly estimate the channel matrix $\mathbf{H}$.

ML and MMSE based detectors are a conventionally used detectors for MIMO systems. The complexity order of the MMSE detector is $O(K^2M^2(n_r + KM) + KM|\mathbb{S}_{\text{SU-MBM}}|)$. The order of complexity for ML detection is $O(n_r KM|\mathbb{S}_{\text{SU-MBM}}|^K)$.

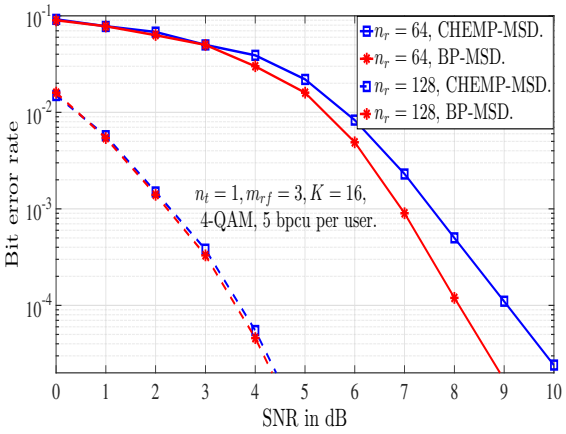


Fig. 9: BER performance of BP-MSD and CHEMP-MSD for multiuser MBM system with $K = 16$, $n_t = 1$, $m_{rf} = 3$, 4-QAM, 5 bpcu per user, and $n_r = 64, 128$.

Therefore, both the proposed message passing algorithms have much lesser computational complexity than the MMSE and ML detectors for large-scale MU-MBM signal detection.

E. Performance results and discussions

In this subsection, we present the BER performance of the proposed BP and CHEMP algorithms in MU-MBM systems. We also compare it with the performance of other MU-MBM detectors discussed before.

Performance of the message passing based detection algorithms: We simulate an MU-MBM system with $n_t = 1, m_{rf} = 3, K = 16$, 4-QAM, 5 bpcu per user, and $n_r = 64, 128$. Note that for $n_r = 64$, the MU-MBM system becomes an under-determined system as the transmit vector $\mathbf{x}$ is of dimension 128×1. For $n_r = 128$, the system is fully-determined.

In Fig. 9, we present the performance of the proposed BP-MSD and CHEMP-MSD algorithms when perfect channel knowledge is available at the BS. From Fig. 9, it can be seen that both the detectors have almost the same BER performance for $n_r = 128$. For $n_r = 64$, BP-MSD performs better than CHEMP-MSD by about 1 dB at 10^{-4} BER. This is because, for under-determined systems, the $\frac{\mathbf{H}^H\mathbf{H}}{n_r}$ matrix will have relatively strong off-diagonal elements. Therefore, the channel hardening effect is less pronounced.

Figure 10 shows the performance of the BP and CHEMP receivers (i.e., with estimated channel knowledge at the BS). From Fig. 10, it can be seen that the BP receiver performs better than the CHEMP receiver by about 1 dB at 10^{-4} BER for $n_r = 64$. This is because, as seen before, the system is under-determined for $n_r = 64$. For $n_r = 128$, the CHEMP receiver outperforms the BP receiver by about 0.8 dB at 10^{-4} BER. This is because, when the MBM channel is fully-determined, directly estimating $\mathbf{H}^H\mathbf{H}$ in CHEMP without explicitly estimating $\mathbf{H}$ improves the detection performance. This shows that the BP receiver algorithm is efficient for under-determined MBM channels and the CHEMP receiver algorithm is efficient for fully- and over-determined MBM channels.

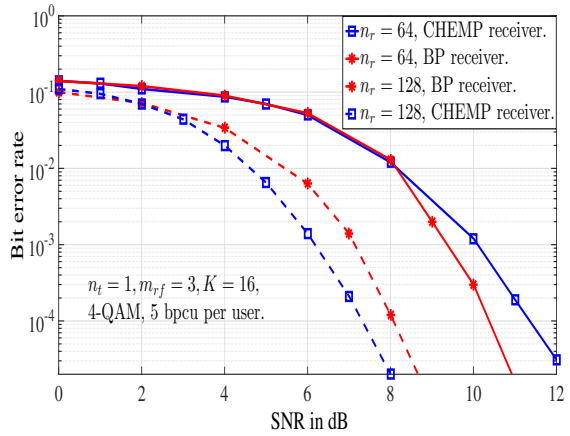


Fig. 10: BER performance of BP receiver and CHEMP receiver for multiuser MBM system with $K = 16$, $n_t = 1$, $m_{rf} = 3$, 4-QAM, 5 bpcu per user, and $n_r = 64, 128$.

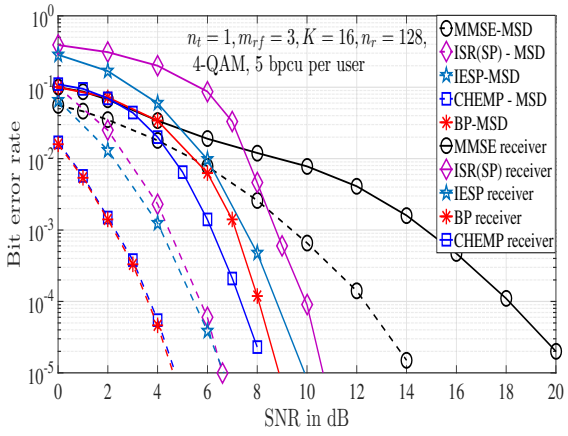
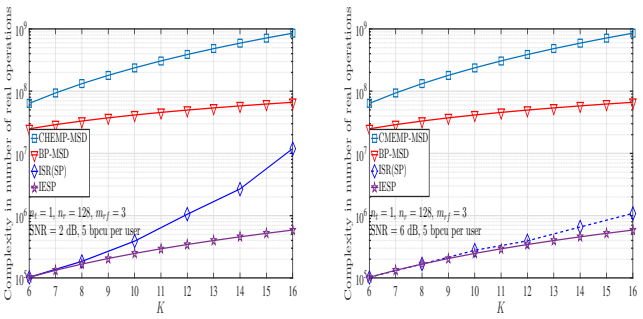


Fig. 11: BER performance comparison between $i)$ MMSE receiver (MMSE detector + MMSE channel estimator), $ii)$ ISR(SP) receiver (ISR(SP) detector + MMSE channel estimator), $iii)$ IESP receiver (IESP detector + MMSE channel estimator), $iv)$ BP receiver, and $v)$ CHEMP receiver for multiuser MBM system with $K = 16$, $n_t = 1$, $m_{rf} = 3$, 4-QAM, and 5 bpcu per user, and $n_r = 128$.

Comparison with other receiver algorithms: Here, we compare the BER performance of the BP and CHEMP receivers with that of the ISR receiver (Sec. III-C), IESP receiver (Sec. III-D), and MMSE receiver. In Fig. 11, we present the performance comparisons between (i) MMSE detector (perfect CSI) and MMSE receiver (MMSE detector with MMSE channel estimator), (ii) ISR(SP) detector (perfect CSI) and ISR(SP) receiver (ISR-SP detector with MMSE channel estimator), $iii)$ IESP detector (perfect CSI) and IESP receiver (IESP detector with MMSE channel estimator), (iv) BP-MSD and BP receiver, and (v) CHEMP-MSD and CHEMP receiver. The MU-MBM system configuration chosen is $n_t = 1, m_{rf} = 3, K = 16, n_r = 128$, 4-QAM, and 5 bpcu per user. From Fig. 11, it can be seen that for both perfect channel knowledge and estimated channel knowledge, the proposed BP and CHEMP algorithms outperform MMSE, ISR, and IESP based algorithms. For instance, at a BER of 10^{-4}, BP-MSD and CHEMP-MSD outperform MMSE detector by about 8.5 dB, ISR(SP) detector by about 2 dB, and IESP detector by about 1.5 dB. For the same BER, the BP receiver outperforms the MMSE receiver by about 10 dB, the ISR(SP) receiver by about 2 dB, and IESP detector by about 1 dB. The CHEMP receiver outperforms the MMSE receiver by about 11 dB, ISR(SP) receiver by about 3 dB, and IESP detector by about 2 dB at a BER of 10^{-4}.

Computational complexity comparison between BP-MSD, CHEMP-MSD, ISR(SP), and IESP:
In Fig. 12, we present a computational complexity comparison between the BP-MSD, CHEMP-MSD, ISR(SP), and IESP detectors obtained through simulations. The simulations are performed for SNR values of 2 dB and 6 dB, and varying number of users. It can be observed that ISR(SP) and IESP algorithms have lesser complexity compared to BP-MSD and CHEMP-MSD algorithms. It can be further observed from Fig. 12(a) that the complexity of ISR(SP) detector increases polynomially with the number of users at a lower SNR value of 2 dB, whereas Fig. 12(b) shows that the complexity of ISR(SP) increases almost linearly with the number of users at a higher



(a) SNR = 2 dB (b) SNR = 6 dB

Fig. 12: Computational complexities of CHEMP-MSD, BP-MSD, ISR(SP), and IESP detectors in number of real operations as a function of K for $n_t = 1$, $n_r = 128$, $m_{rf} = 3$, and 5 bpcu per user. (a) SNR = 2 dB, (b) SNR = 6 dB.

SNR value of 6 dB. For all values of SNR, the IESP detector has the lowest complexity compared to the other considered detectors. From Figs. (11) and (12), we can observe that message passing detectors achieve better BER performance at the expense of higher computational complexity.

V. Conclusion

MBM is a recent and attractive modulation scheme that employs RF mirrors (parasitic elements) as controlled scatterers near the transmit antenna to create the channel alphabet so that additional information bits are conveyed through indexing of these mirrors. In the literature, MBM has been shown to offer RF hardware complexity and performance advantages in point-to-point single user communications. Here, we investigated the use of MBM in multiuser communication on the uplink. We showed that MBM can offer significant performance gains in multiuser settings as well. Our results showed that MU-MBM significantly outperforms MU-CM and MU-SM, which

are other popularly known systems with users having single transmit RF chain. We also proposed detection algorithms based on compressive sensing and message passing techniques. An efficient channel estimation scheme that exploited the channel hardening effect was also proposed. The results of the proposed detectors/receivers demonstrated that MBM is an attractive modulation scheme suited for use in large-scale multiuser communications on the uplink.

APPENDIX A
ASYMPTOTIC CAPACITY OF MBM

Consider an MBM system with a single transmit antenna and m_{rf} RF mirrors placed near it. Let n_r be the number of receive antennas. The received MBM signal vector is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{37}$$
$$= \mathbf{H}\mathbf{m}x + \mathbf{n}, \tag{38}$$

where $\mathbf{y} \in \mathbb{C}^{n_r \times 1}$ is the received signal vector, $\mathbf{x} \in \mathbb{S}_{\text{SU-MBM}}$ is the $M \times 1$ transmit MBM signal vector, $\mathbb{S}_{\text{SU-MBM}}$ is the single-user MBM signal set defined in (2), $\mathbf{n} \in \mathbb{C}^{n_r \times 1}$ is the noise vector distributed $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_{n_r})$, and $\mathbf{H} \in \mathbb{C}^{n_r \times M}$ is the channel matrix with ith column being the $n_r \times 1$ channel vector corresponding to the ith MAP. The entries of $\mathbf{H}$ are i.i.d $\mathcal{CN}(0,1)$. The transmit signal vector can be written as $\mathbf{x} = \mathbf{m}x$, where $\mathbf{m}$ is the MAP vector; if ith MAP is selected, then $m_i = 1$ and $m_j = 0,\ \forall\ j \neq i$. Further, x is the conventional source modulation symbol (e.g. QAM/PSK). In the case of pure MBM (without source modulation), $x = 1$ (tone). For the rest of this Appendix, we consider an MBM system with a tone. Thus, if ith MAP is selected for transmission in a channel use, then the the received signal vector is given by

$$\mathbf{y} = \mathbf{H}\mathbf{m}_i + \mathbf{n} = \mathbf{h}_i + \mathbf{n} \tag{39}$$

where $\mathbf{h}_i$ is the ith column of $\mathbf{H}$, which is the channel vector corresponding to the ith MAP, with $\mathbf{m}_i$ denoting the ith MAP. The capacity of this MBM system can be written as

$$\begin{aligned} C &= \mathbb{E}_{\mathbf{H}}(h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x})) \\ &= \mathbb{E}_{\mathbf{H}}(h(\mathbf{y}) - h(\mathbf{n})) \\ &= \mathbb{E}_{\mathbf{H}}(h(\mathbf{y}) - \log_2 \det(\pi e \sigma^2 \mathbf{I}_{N_r})), \end{aligned} \tag{40}$$

where $h(\cdot)$ denotes the differential entropy. We first consider the case of one receive antenna and then extend it to the case of n_r receive antennas. If ith MAP is selected, with the knowledge of $\mathbf{H}$, the received signal becomes

$$y = h_i + n_i, \tag{41}$$

where $n_i \sim \mathcal{CN}(0, \sigma^2)$. Therefore,

$$f(y|\mathbf{H}, \mathbf{m}_i) = f(y|h_i) = \frac{1}{\pi\sigma^2} e^{-|y-h_i|^2/\sigma^2}. \tag{42}$$

Since all the M MAPs are equally likely

$$f(y|\mathbf{H}) = \frac{1}{M}\sum_{i=1}^{M} f(y|\mathbf{H}, \mathbf{m}_i) \tag{43}$$
$$= \frac{1}{M}\sum_{i=1}^{M} \frac{1}{\pi\sigma^2} e^{-|y-h_i|^2/\sigma^2}. \tag{44}$$

This equation can not be simplified further. However, in the case when $M \to \infty$ (i.e, $m_{rf} \to \infty$), every possible value of h_i occurs according to its distribution. The received signal y can then be expressed as [23]

$$f(y|\mathbf{H}) = f_h(h) * f_n(n) \tag{45}$$
$$= \int_T f_h(\tau) f_n(y-\tau) d\tau \tag{46}$$
$$= \int_T \frac{1}{\pi} e^{-|y-\tau|^2} \frac{1}{\pi\sigma^2} e^{-\frac{|\tau|^2}{\sigma^2}} d\tau \tag{47}$$
$$= \frac{1}{\pi(1+\sigma^2)} e^{-\frac{|y|^2}{\sigma^2+1}}, \tag{48}$$

where T denotes the support set of τ, which is the entire complex plane. Equation (45) follows from the fact that h and n are independent R.Vs and y is a sum of independent R.Vs. Hence, distribution of y is the convolution of distributions of h and n. The differential entropy of the received signal is then given by

$$h(y) = -\int \frac{1}{\pi(1+\sigma^2)} e^{-\frac{|y|^2}{\sigma^2+1}} \log_2 \left(\frac{1}{\pi(\sigma^2+1)} e^{-\frac{|y|^2}{\sigma^2+1}} \right) dy \tag{49}$$
$$= \log_2 \left[\pi e(\sigma^2+1)\right]. \tag{50}$$

Therefore, the capacity of MBM system with $m_{rf} \to \infty$ and single receive antenna is given by

$$C = \log_2 \left[\pi e(\sigma^2+1)\right] - \log_2(\pi e \sigma^2) \tag{51}$$
$$= \log_2 \left(1 + \frac{1}{\sigma^2}\right) \tag{52}$$
$$= \log_2(1+\gamma), \tag{53}$$

where γ is the SNR. From (53), we see that the capacity of MBM system with single receive antenna, as $m_{rf} \to \infty$, is same as that of SISO AWGN channel capacity.

Next, we consider the case of n_r receive antennas. Since $y_1, y_2, \cdots, y_{n_r}$, the received signals across n_r Rx antennas, are independent and uncorrelated, the joint pdf of received signal $\mathbf{y}$ is just the product of pdfs of received signals across n_r receive antennas. Therefore, the pdf of $\mathbf{y}$ as $m_{rf} \to \infty$, is given by

$$f(\mathbf{y}) = \frac{1}{[\pi(1+\sigma^2)]^{n_r}} e^{-\frac{\|\mathbf{y}\|^2}{1+\sigma^2}}, \tag{54}$$

where $\mathbf{y} = [y_1 y_2 \cdots y_{n_r}]^T$ is the received signal vector. It also follows from the independence of $y_1, y_2, \cdots, y_{n_r}$ that, the joint entropy is just the sum of their individual entropies. Hence, the capacity of MBM system with n_r receive antennas, as $m_{rf} \to \infty$, is given by

$$C = n_r \log_2[\pi e(1+\sigma^2)] - n_r \log_2(\pi e \sigma^2) \tag{55}$$
$$= n_r \log_2 \left(1 + \frac{1}{\sigma^2}\right) \tag{56}$$
$$= n_r \log_2(1+\gamma). \tag{57}$$

From (57), it can be seen that the capacity of an MBM system with n_r receive antennas asymptotically (as $m_{rf} \to \infty$) achieves the capacity of n_r parallel AWGN channels.

REFERENCES

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge Univ. Press, 2005.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186-195, Feb. 2014.

[3] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998-1009, Apr. 2016.

[4] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653-656, Oct. 2014.

[5] E. Basar, M. Wen, R. Mesleh, M. D. Renzo, Y. Xiao, and H. Haas, "Index modulation techniques for next-generation wireless networks, *IEEE Access*, vol. 5, pp. 16693-16746, Sep. 2017.

[6] R. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Tech.*, vol. 57, no. 4, pp. 2228-2241, Jul. 2008.

[7] M. D. Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, opportunities and implementation," *Proceedings of the IEEE*, vol. 102, no. 1, pp. 56-103, Jan. 2014.

[8] A. Chockalingam and B. S. Rajan, *Large MIMO Systems*, Cambridge Univ. Press, Feb. 2014.

[9] M. A. Sedaghat, V. I. Barousis, R. R. Muller, and C. B. Papadias, "Load modulated arrays: a low-complexity antenna," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 46-52, Mar. 2016.

[10] M. A. Sedaghat, R. R. Muller, G. Fischer, and A. Ali, "Discrete load-modulated single-RF MIMO transmitters," *Proc. WSA2016*, Mar. 2016

[11] S. Bhat and A. Chockalingam, "Detection of load-modulated multiuser MIMO signals," *IEEE Wireless Comm. Lett.*, 2017. doi: 10.1109/LWC.2017.2771312.

[12] A. K. Khandani, "Media-based modulation: a new approach to wireless transmission," *Proc. IEEE ISIT'2013*, pp. 3050-3054, Jul. 2013.

[13] A. K. Khandani, "Media-based modulation: converting static Rayleigh fading to AWGN," *Proc. IEEE ISIT'2014*, pp. 1549-1553, Jun.-Jul. 2014.

[14] E. Seifi, M. Atamanesh, and A. K. Khandani, "Media-based modulation: a new frontier in wireless communications," online arXiv:1507.07516v3 [cs.IT] 7 Oct. 2015.

[15] Y. Naresh and A. Chockalingam, "On media-based modulation using RF mirrors," *IEEE Trans. Veh. Tech.*, vol. 66, no. 6, pp. 4967-4983, Jun. 2017.

[16] E. Seifi, M. Atamanesh, A. K. Khandani, "Media-based MIMO: outperforming known limits in wireless," *Proc. IEEE ICC'2016*, pp. 1-7, May 2016.

[17] O. N. Alrabadi, A. Kalis, C. B. Papadias, R. Prasad, "Aerial modulation for high order PSK transmission schemes," *Proc. Wireless VITAE 2009*, pp. 823-826, May 2009.

[18] O. N. Alrabadi, A. Kalis, C. B. Papadias, and R. Prasad, "A universal encoding scheme for MIMO transmission using a single active element for PSK modulation schemes," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5133-5142, Oct. 2009.

[19] M. D. Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: challenges, opportunities and implementation," *Proc. of the IEEE*, vol. 102, no. 1, pp. 56-103, Jan. 2014.

[20] A. Chockalingam and B. S. Rajan, *Large MIMO Systems*, Cambridge Univ. Press, Feb. 2014.

[21] T. L. Narasimhan and A. Chockalingam, "On the capacity and performance of generalized spatial modulation," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 252-255, Feb. 2016.

[22] T. L. Narasimhan, P. Raviteja, and A. Chockalingam, "Generalized spatial modulation in large-scale multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3764-3779, Jul. 2015.

[23] H. Yonghong, W. Pichao, W. Xiang, Z. Xiaoming, and H. Chunping, "Ergodic capacity analysis of spatially modulated systems," *China Commun.*, vol. 10, no. 7, pp. 118-125, Jul. 2013.

[24] S. Lokya, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Comm. Lett.,* vol. 5, no. 9, pp. 369-371, Sep. 2001.

[25] Z. Gao, L. Dai, S. Han, C-L. I, Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.,* pp. 2-11, 2018. doi: 10.1109/MWC.2017.1700147.

[26] G. Wunder, H. Boche, T. Strohmer, and P. Jung, "Sparse signal processing concepts for efficient 5G system design," *IEEE Access,* vol. 3, pp. 195-208, Feb. 2015.

[27] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, vol. 12, pp. 4655-4666, Dec. 2007.

[28] D. Needell and J. A. Tropp, "CoSaMP: iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis* 26.3 (2009): 301-321.

[29] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 2230-2249, May 2009.

[30] A. G. Rodriguez and C. Masouros, "Low-complexity compressive sensing detection for spatial modulation in large-scale multiple access channels," *IEEE Trans. Commun.*, vol. 63, no. 7, pp. 2565-2579, Jul. 2015.

[31] Z. Gao, L. Dai, Z. Wang, S. Chen, and L. Hanzo, "Compressive-sensing-based multiuser detector for the large-scale SM-MIMO uplink," *IEEE Trans. on Veh. Tech.*, vol. 65, no. 10, pp. 8725-8730, Oct. 2016.

[32] B. Shamasundar and A. Chockalingam, "Multiuser media-based modulation for massive MIMO Systems," *Proc. IEEE SPAWC'2017*, Jul. 2017.

[33] T. L. Narasimhan and A. Chockalingam, "Channel hardening-exploiting message passing (CHEMP) receiver in large-scale MIMO systems," *IEEE J. Sel. Topics in Signal Process*., vol. 8, no. 5, pp. 847-860, Oct. 2014.

[34] S. Jacob, T. L. Narasimhan, and A. Chockalingam, "Message passing receivers for large-scale multiuser media-based modulation," accepted in *IEEE PIMRC'2017*, Oct. 2017.

[35] A. Bandi and C. R. Murthy, "Structured sparse recovery algorithms for data decoding in media based modulation," *Proc. IEEE ICC'2017* , Jul. 2017.

[36] E. J. Candes and T. Tao, "Near-optimal signal recovery from projections: universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406-5425, Dec. 2006.

[37] M. Pretti, "A message passing algorithm with damping," *J. Stat. Mech.: Theory and Practice*, Nov. 2005. doi:10.1088/1742-5468/2005/11/P11008.