Received May 17, 2019, accepted June 13, 2019, date of publication June 17, 2019, date of current version July 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923445

Coherent and Non-Coherent Multilayer Index Modulation

SARATH GOPI, SHEETAL KALYANI, AND LAJOS HANZO

Corresponding author: Lajos Hanzo (lh@ecs.soton.ac.uk)

The work of L. Hanzo was supported in part by the Engineering and Physical Sciences Research Council under Project EP/Noo4558/1 and Project EP/PO34284/1, and in part by the COALESCE of the Royal Society's Global Challenges Research Fund Grant, and in part by the European Research Council's Advanced Fellow Grant QuantCom.

ABSTRACT Index modulation assisted OFDM (OFDM-IM) has become a popular technique in wireless communications. In this paper, we proposes a novel multilayer framework for OFDM-IM. Being quite generic in its nature, the framework can be amalgamated with most of the OFDM-IM variants. For coherent OFDM-IM, we propose three different schemes; 1) multilayer index modulation (MIM); 2) dual mode MIM; and 3) compressive sampling aided MIM. The schemes strikes a compelling tradeoff between the BER, energy efficiency and throughput. Finally, we extend this multilayer framework to non-coherent index modulation, where the structure provides an improved BER performance compared to its single layer counterpart, albeit this is achieved at the cost of reduced energy efficiency. The performance analysis of the proposed scheme is confirmed by our simulation results.

INDEX TERMS Compressed sensing, energy efficiency, index modulation, non-coherent modulation, OFDM.

LIST OF ACRONYMS

AWGN	Additive white Gaussian noise
BER	Bit error rate
CLER	Cluster error rate
CSMIM	Compressed sampling aided MIM
DMIM	Dual mode MIM
E_b	Energy required per bit transmission
FD	Frequency domain
ICI	Inter carrier interference
IER	Index error rate
LLR	Log likelihood ratio
MIM	Multilayer index modulation
NC	Non-coherent
OFDM-IM	OFDM with index modulation
PAPR	Peak-to-average power ratio
RIP	Restricted isometric property
SNR	Signal to noise ratio
TD	Time domain
TP	Throughput (Number of bits transmitted) per OFDM frame

LIST OF SYMBOLS

$\mathbf{A}^c$	Complementary set of $\mathbf{A}$
$\|.\|^p$	p^{th} norm of the vector
$\|.\|$	Cardinality of the set
$\lfloor . \rfloor$	Floor of the number
x	Transmitted OFDM frame in frequency domain
x_T	Transmitted OFDM frame in time domain
y	Received OFDM frame in frequency domain
y_T	Received OFDM frame in time domain
h	Channel frequency response
h_T	Channel impulse response
w_F	Noise in frequency domain
w_T	Noise in time domain
σ^2	Noise variance in frequency domain
σ_T^2	Noise variance in time domain
σ_h^2	Variance of channel frequency response
$\mathcal{CN}(\mu, C)$	Complex Gaussian distribution with mean μ and covariance matrix C
s_i	A symbol from some constellation
Pr(.)	Probability of some event
F_p	Partial Fourier matrix
$P_{\mathtt{ce}}$	Probability of cluster error
$P_{\mathtt{ie}}$	Probability of index error
Γ_b	Set of sub-carrier indices having non-zero values
$\mathcal{I}_b$	$\|\Gamma_b\|$
$\mathcal{B}(n_1, n_2)$	Beta function
$\mathcal{I}_\alpha(n_1, n_2)$	Regularized incomplete Beta function

The associate editor coordinating the review of this manuscript and approving it for publication was Md Fazlul Kader.

I. INTRODUCTION

In recent years, the field of wireless communication has witnessed the emergence of a large number of variants of the classic OFDM system relying on index modulation (OFDM-IM) [1]–[4]. Explicitly, in OFDM-IM the information is conveyed both by classic symbols, as well as by the specific subcarrier patterns it uses for transmitting the classic symbols. It may be deemed to be one of the most promising techniques for next generation wireless communication. The popularity of this approach is due to its low BER, improved peak-to-average power ratio (PAPR), robustness to inter carrier interference (ICI) and flexibility in selecting the number of activated subcarriers according to the required trade-off between its bandwidth efficiency and bit error rate (BER) [2].

OFDM-IM concept was introduced in the late nineties by Frenger and Svensson [5]. However, Ba ar *et al.* [6] have recently rekindled the research community's interest. To elaborate, in OFDM-IM, each OFDM block of N subcarriers is split into B clusters each having L subcarriers, so that $N = LB$. In each of these clusters, only K out of L subcarriers are used for transmitting classic symbols. Additionally, a part of the incoming bit sequence is used for selecting these K subcarriers in each of the clusters. Then, the classic symbols are mapped to the appropriately selected active subcarriers. Hence, in OFDM-IM, the information is conveyed not only by the classic PSK/QAM symbols, but also by the activated pattern of subcarriers that convey the classic symbols.

However, a common criticism of OFDM-IM is its reduced spectral efficiency compared to classic OFDM, since a part of the subcarriers remains unused, albeit the power conserved by the passive subcarriers may be assigned to the active ones. Hence, a number of variants have been reported in the literature for improving the performance. For example, in the dual mode index modulation [7], a twin set of modulating symbols is transmitted by activating the unused subcarriers of the original OFDM-IM, so that the throughput (TP) is improved. In another variant of OFDM-IM, namely in quadrature index modulation [8], two separate bit sequences are used for selecting the in-phase and the quadrature component of the activated subcarriers. In [9], a pair of information guided pre-coding schemes have been adopted for improving the throughput, while a Huffman coded subcarrier selection procedure is used in [10]. A pair of enhanced OFDM-IM versions have been reported in [11], where one of the methods uses a joint I and Q subcarrier index selection, while the other advocates a linear constellation pre-coding technique for achieving diversity gain. A coordinated interleaved OFDM-IM technique was proposed in [12] for combating block fading and a trellis coded modulation scheme was designed for improving the detection performance of OFDM-IM [13]. Index modulation techniques have also been adopted in vector OFDM [14], DCT OFDM [15], filter bank multi-carrier systems [16], in generalized frequency division multiplexing [17], [18], generalized space and frequency modulation [19], in MIMO OFDM [20]–[22], non-coherent communication [23] and so on. A detailed review of OFDM-IM and its variants is provided in [2] and [24].

Against this backdrop, we propose an improved OFDM-IM scheme namely 'Multilayer Index Modulation' (MIM). The idea is to make a second level cluster selection for conveying additional input bits. Hence, the information is not only transmitted by the classic PSK/QAM symbols and the activated subcarrier locations that carry the symbols, but also by the set of information-dependent clusters selected. The principle of multi layer modulation based on OFDM is not entirely new in the literature [25]–[27]. Their unifying principle is that additional PSK/QAM symbols are incorporated into the multi layer structure for enhancing its throughput. By contrast, our proposed scheme does not introduce any extra PSK/QAM symbols into the multi layer structure. Instead, the multi layer architecture itself acts as an information bearing entity. Since no extra PSK/QAM symbols are used by our MIM scheme, there will be no additional transmit energy dissipation. The MIM scheme is generic in the sense that most of the other index modulation techniques can be combined with this scheme for exploiting its advantages. The main contributions of the paper are the following.

- The new MIM philosophy is proposed, which requires less energy per bit than the conventional OFDM-IM system, albeit at the cost of a potentially reduced throughput. Therefore, in order to improve its throughput, we propose two further variants of the proposed scheme.
  - In the first variant, we extend the concept of dual mode IM [7] to improve the throughput. The resultant Dual Mode MIM (DMIM) uses symbols from a different constellation to fill the unselected clusters.
  - The second approach, compressive sampling aided MIM (CSMIM) directly transforms the OFDM-IM data to its sub-sampled time series using a partial Fourier matrix and thereby improving the throughput.

  Furthermore, we derive low complexity detectors for each of the proposed schemes and study their performance by simulations.
- We also develop a non-coherent MIM scheme as an extension of non-coherent index modulation [23]. Our proposed non-coherent MIM system provides better throughput and BER performance than its single layer counterpart [23] at the cost of reduced energy efficiency.
- We derive the probability of error in the non-coherent MIM scheme and evaluates its performance by simulations.

The proposed MIM philosophy can be applied recursively, which forms a multi layer architecture for our index modulation scheme. The advantage of the multi layer architecture is that as the number of layers increases more information will be embedded into the clusters, rather than into independent subcarriers, which improves the BER performance of the systems as demonstrated later in Section V-A. Hence,

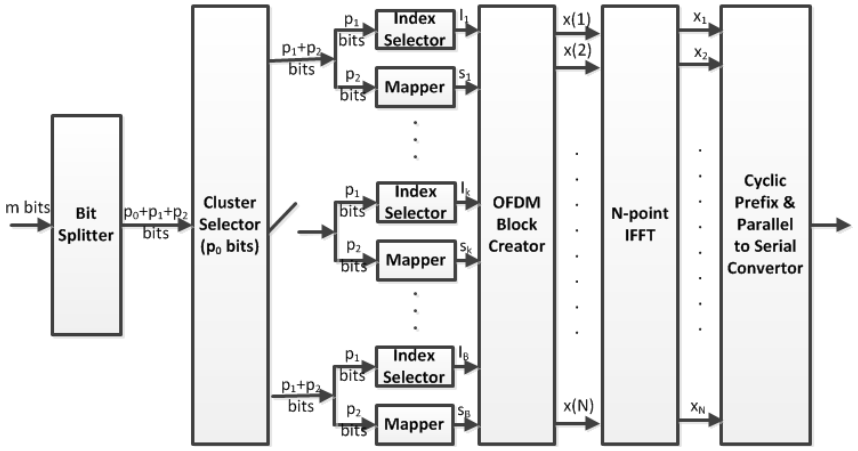

FIGURE 1. Proposed multilayer OFDM-IM (MIM).

the multilayer framework strikes a trade off between the energy efficiency, BER performance and throughput.

The rest of the paper is organized as follows. In Section II, our MIM scheme is detailed. Its variants, DMIM and CSMIM are also introduced in this section. The detectors conceived for the proposed schemes are discussed in Section III. Section IV describes the non-coherent version of the conceived MIM. Finally, our simulation results are discussed in Section V and we conclude in Section VI.

II. THE PROPOSED MULTI LAYER INDEX MODULATION (MIM) SCHEMES

In this section, we will demonstrate the benefits of our MIM schemes. To elucidate the MIM concept, first a brief description of OFDM-IM is provided in the next subsection.

A. OFDM-IM

In the conventional OFDM-IM, the subcarriers in an OFDM frame are partitioned into a number of clusters and in each cluster only a section of subcarriers is used for transmitting PSK/QAM symbols. The information is carried both by the PSK/QAM symbols and by the pattern of subcarriers that convey the PSK/QAM symbols.

Consider an OFDM-IM system bearing a total of N subcarriers, that are partitioned into B clusters with each containing L subcarriers, where we have $N = LB$. Let $\mathbf{x} = \left[\mathbf{x}_0^T, \ldots, \mathbf{x}_{B-1}^T\right]^T$ be an $N \times 1$ vector representing an OFDM frame. Here, $\mathbf{x}_b = [x(bL), x(bL+1), \ldots, x((b+1)L-1)]^T$ is the b^{th} cluster, where $b = 0, 1, \ldots, B-1$. In OFDM-IM, only K out of L subcarriers are activated in a cluster for carrying a non-zero PSK/QAM symbol. The OFDM frame $\mathbf{x}$ is converted into the corresponding time series and N_{CP} cyclic prefix samples are concatenated. In the channel, the transmitted signal is convolved with the channel response. At the receiver, after removing the cyclic prefix, the received data is converted back to the frequency domain (FD) yielding:

$$y(n) = x(n)h(n) + w_F(n), \quad \text{for } n = 1, \ldots, N \qquad (1)$$

where $h(n) \sim \mathcal{CN}(0, 1)$ represents the independent complex channel fading coefficients, while $w_F(n) \sim \mathcal{CN}(0, \sigma^2)$ represents the noise samples in the FD. Let us assume that m bits are transmitted per OFDM frame. Then the throughput of the system is $m/(N + N_{CP})$ bits/s/Hz [6], assuming zero Nyquist filtering excess bandwidth.

B. OFDM-MIM

In the proposed scheme, we introduce a second-level selection procedure. This is achieved by appropriately selecting the clusters to transmit additional data, i.e., instead of modulating K subcarriers of all the clusters, only Q of the B clusters actually convey PSK/QAM symbols. This set of Q clusters can be selected from the B clusters in $\binom{B}{Q}$ ways and the specific selection of these Q clusters is based on the incoming bit sequence. Hence, in this scheme, the information is carried by 1) the specifically selected pattern of clusters 2) the information-dependent choice of K out L subcarriers in each of the selected clusters and 3) the PSK/QAM symbols on the activated subcarriers. Therefore, the incoming bit sequence is split into three groups, where the first $p_0 = \left\lfloor \log_2 \binom{B}{Q} \right\rfloor$ bits determine the active or Type I clusters, while the second $p_1 = Q \left\lfloor \log_2 \binom{L}{K} \right\rfloor$ bits choose the active subcarriers in each of the active clusters and the final $p_2 = QK \log_2 M$ bits decide PSK/QAM symbols. The total number of bits is $p = p_0 + p_1 + p_2$. The block diagram of the proposed scheme is shown in Fig. 1

Our MIM principle can also be applied recursively. Explicitly, instead of activating K subcarriers in a cluster, the cluster itself can be split into different sub-clusters for a second layer selection, which can be further repeated in the next level. This forms a multi-layer architecture, in which selections can be carried out at each layer depending on the input bit sequence.

To compare the energy efficiency of the MIM scheme, let us assume that E is the energy required for transmitting a message symbol in a single subcarrier. Then the energy per

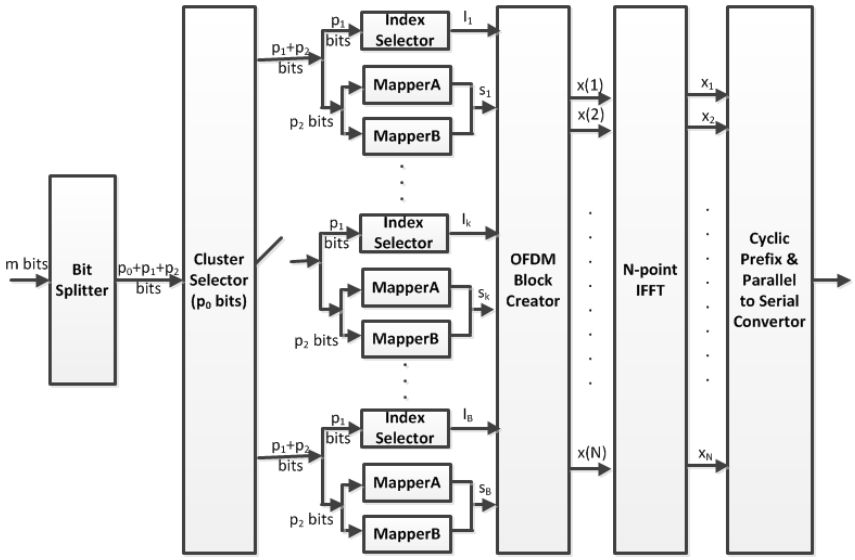

FIGURE 2. Proposed dual mode multilayer OFDM-IM (DMIM).

bit of a transmission is

$$E_b = \frac{E}{\frac{1}{K}\lfloor \log_2 \binom{L}{K} \rfloor + \log_2 M + \frac{1}{KQ}\lfloor \log_2 \binom{B}{Q} \rfloor}, \quad (2)$$

whereas the corresponding energy requirement in the case of OFDM-IM is $E_b = \frac{E}{\frac{1}{K}\lfloor \log_2 \binom{L}{K} \rfloor + \log_2 M}$, which is bigger than that of MIM, since the denominator of (2) has an additional term.

However, the throughput becomes potentially lower than that of OFDM-IM, since $(B - Q)$ clusters are idle. Explicitly, the throughput of the proposed scheme is $\frac{Q\lfloor \log_2 \binom{L}{K} \rfloor + QK \log_2 M + \lfloor \log_2 \binom{B}{Q} \rfloor}{N+N_{CP}}$, whereas it is $\frac{B\lfloor \log_2 \binom{L}{K} \rfloor + BK \log_2 M}{N+N_{CP}}$ in the case of OFDM-IM. In order to circumvent this loss, we propose the two techniques in the next subsection

C. MIM VARIANTS

1) DUAL MODE MIM SCHEME (DMIM)

In this scheme, we further invoke the dual-mode concept of [7]. More specifically, instead of keeping $(B - Q)$ clusters idle, they can be modulated using PSK/QAM symbols from a second constellation. The block diagram of this scheme is shown in Fig. 2. In this case, the throughput becomes $\frac{B\lfloor \log_2 \binom{L}{K} \rfloor + BK \log_2 M + \lfloor \log_2 \binom{B}{Q} \rfloor}{N+N_{CP}}$, which is higher than that of OFDM-IM. Here, the total energy required is BKE, which is the same as that of OFDM-IM. However, DMIM transmits an additional $\left\lfloor \log_2 \binom{B}{Q} \right\rfloor$ number of bits. Hence, the energy per bit is reduced by this scheme and its throughput is increased.

2) COMPRESSIVE SAMPLING AIDED MIM (CSMIM)

In our original MIM scheme, only Q of B clusters are active and in each of these Q clusters, only K subcarriers have non-zero symbols. Hence, only KQ of LB subcarriers have non-zero symbols. For small values of K and Q, our OFDM-MIM frame becomes sparse in the FD. Hence, compressive sampling can be used for transmitting a sub-sampled time series of the data. Explicitly, we use a partial Fourier matrix for directly transforming the OFDM-IM data to its sub-sampled time series. The partial Fourier matrix is constructed by retaining a set of random rows of the Fourier matrix. In [28], a compressed sampling based technique has been suggested for OFDM-IM, where a random matrix is used to first compress the OFDM-IM data in the FD. The compressed data is then transformed to the time domain (TD) and transmitted as in conventional OFDM. However, in our case, the compression and the transformation to the TD take place in a single step. Therefore, the computational complexity of the proposed scheme is reduced compared to that of [28]. At the receiver, suitable compressed sensing techniques can be adopted for recovering the actual data from the sub-sampled time series.

The diagrammatic representation of the scheme is almost the same as that of Fig. 1, except for the IFFT block that transforms the FD data to time series. In this scheme, instead of using the classic Fourier matrix to transform the data to the TD, we propose to use a partial Fourier matrix. The idea behind using a partial Fourier matrix is to reduce the frame size to be transmitted. Finally, the usual cyclic prefix is concatenated with the sub-sampled time series and transmitted. In this case, the throughput will become $\frac{B\lfloor \log_2 \binom{L}{K} \rfloor + BK \log_2 M + \lfloor \log_2 \binom{B}{Q} \rfloor}{N_R+N_{CP}}$, where $N_R < N$ is the number of rows in the partial Fourier matrix used for generating the time series. It is evident that the throughput of this variant is better than that of the MIM, while requiring the same energy for transmission.

The encoding schemes of conventional OFDM-IM are directly applicable to the proposed schemes. However, in our

case, the incoming bit sequence is split into three groups, whereas only into two groups for conventional OFDM-IM. Explicitly, p_0 and p_1 bits can be used to identify the active clusters and active subcarriers, respectively. This can be achieved using either a look up table based method or the combinatoric approach of [6]. Finally, p_2 bits are conveyed using M-ary PSK/QAM symbols, which are mapped to the activated subcarriers of the active clusters. The detectors conceived for the proposed schemes are discussed in the next section.

III. DETECTORS

In the proposed schemes, we have to detect the active or Type I clusters, the active subcarriers of the active clusters and the PSK/QAM symbols of the active subcarriers. Hence, the optimal detector is a three dimensional joint ML detector. This detector is however impractical owing to its high computational complexity, which is on the order of $\mathcal{O}(c_0 c_1 M^{QK})$, where $c_0 = 2^{p_0}$ and $c_1 = 2^{p_1}$. Therefore, a Log Likelihood Ratio (LLR) based detector is preferred. The LLR detectors of the MIM and DMIM schemes are similar to each other, whereas the CSMIM detector is different, hence it is discussed in Subsection III-C.

A. LLR DETECTOR FOR MIM

The LLR detection is carried out in three steps. The first step is to detect the active clusters, followed by the detection of active subcarriers. In the last step the PSK/QAM symbols on the active subcarriers in each of the active clusters are detected.

1) ACTIVE CLUSTER DETECTION:

The likelihood ratio of detecting an active cluster is

$$\gamma_b = \sum_{n=bL}^{(b+1)L-1} \log\left(\sum_{s_i\in\mathcal{M}} e^{-\frac{\|s_i h\|^2 + 2\mathbf{Re}(s_i h(n) y(n))}{2\sigma^2}} + \frac{M(L-K)}{K}\right) + \log\left(\frac{QK}{(B-Q)M}\right), \quad (3)$$

where $\mathcal{M}$ is the set of PSK/QAM constellation symbols and $M = |\mathcal{M}|$. The derivation of (3) is given in Appendix (A). Clearly, $\gamma_b > 0$ indicates that the b^{th} cluster is active. However, we know that there are exactly Q active clusters. Therefore, we can calculate the γ_b values for all the clusters, i.e., $b = 0, \ldots, B-1$ and can select that particular set of Q clusters, which has the maximum γ_b values, i.e, $\hat{B}_Q = \{b(1)\ b(2)\ ,\ \ldots,\ b(Q)\}$, where $\gamma_{b(1)} \geq \gamma_{b(2)} \geq \cdots \geq \gamma_{b(Q)} \geq \ldots \geq \gamma_{b(B)}$. Observe that while deriving (3), it is assumed that $Q < B$. When $Q = B$, no selection is possible in the clusters, hence the scheme reduces to conventional OFDM-IM. Subsequently, the active subcarrier and symbol detection can be carried out in these active clusters, for which the procedure is detailed below.

2) LLR FOR ACTIVE SUBCARRIER AND SYMBOL DETECTION

Both the active subcarrier detection and PSK/QAM symbol detection can be carried out similarly to the OFDM-IM system, except that it is applied to the active clusters only. The LLRs of active subcarrier detection are

$$\gamma_n = \log\left(\frac{\sum_{s_i\in\mathcal{M}} \Pr(x(n) = s_i | y(n))}{\Pr(x(n) = 0 | y(n))}\right). \quad (4)$$

Note that any of the PSK/QAM symbols can be mapped to a subcarrier, this is why we have the summation in the numerator of (4). This ratio will turn out to be [2, (12)]

$$\gamma_n = \log K - \log(L-K) + \frac{\|y(n)\|^2}{\sigma^2} - \log\left(\sum_{i=1}^{M}\left(-\frac{1}{\sigma^2}\|y(n) - h(n)s_i\|^2\right)\right).$$

Similarly to the active cluster detection, we know that there are exactly K active subcarriers. Therefore, we can calculate the γ_n values for all subcarriers in an active cluster and select that particular set of subcarriers, which has the largest γ_n values, i.e, $\hat{A}_K = \{n(1)\ n(2)\ ,\ \ldots,\ n(K)\}$, where $\gamma_{n(1)} \geq \gamma_{n(2)} \geq \cdots \geq \gamma_{n(K)} \geq \cdots \geq \gamma_{n(L)}$.

The set of active clusters and active subcarriers in each of the active clusters is converted to the corresponding bit sequence. This can be carried out by using the inverse logic of converting the bit sequence to the selection of active clusters and subcarriers. Finally, the modulating symbols in each of the active subcarriers are directly estimated as

$$\hat{s}(n) = \min_{s_i\in\mathcal{M}} \|y(n) - s_i h(n)\|. \quad (5)$$

B. LLR DETECTOR FOR DMIM

The detection process is exactly same as that of MIM, except that in DMIM, there are no idle clusters. In the non-active (Type II) clusters, different PSK/QAM symbols are loaded. In this case, the likelihood ratio for the active cluster detection can be found in the same way as for MIM. Here, the active cluster is the one in which PSK/QAM symbols are selected from the first constellation, say $\mathcal{M}_1$. The Log Likelihood Ratio is given by

$$\gamma_b = \sum_{n=BL}^{(b+1)L-1} \log\left(\sum_{s_i\in\mathcal{M}_1} \mathbf{e}^{-\frac{\|s_i h\|^2 + 2\mathbf{Re}(s_i h(n) y(n))}{2\sigma^2}} + \frac{M_1(L-K)}{K}\right) - \sum_{n=bL}^{(b+1)L-1} \log\left(\sum_{s_i\in\mathcal{M}_2} \mathbf{e}^{-\frac{\|s_i h(n)\|^2 + 2\mathbf{Re}(s_i h(n) y(n))}{2\sigma^2}} + \frac{M_2(L-K)}{K}\right) + \log\left(\frac{QM_2}{(B-Q)M_1}\right), \quad (6)$$

where $\mathcal{M}_2$ is the second set of PSK/QAM constellation symbols. The detection of Type I clusters can be carried out in the same way as in the case of MIM. Since we know that there are Q Type I clusters, we chose Q clusters having the largest

γ_b values as the Type I (or active) clusters. In (6), we also assume that $Q < B$.

After detecting the active clusters, the detection of active subcarriers and modulating symbols can be done in the same way as for MIM. However, in each of the clusters, the symbols from the appropriate PSK/QAM constellation have to be used. Explicitly, in the Type I clusters, the PSK/QAM symbols from the first constellation are chosen, whereas in the Type II clusters, that from the second PSK/QAM constellation are used.

C. DETECTOR FOR CSMIM

The signal at the CSMIM receiver (after cyclic prefix removal) can be written as

$$\begin{aligned} \mathbf{y}_T &= \mathbf{h}_T * F_p^H \mathbf{x} + \mathbf{w}_T \\ &= \mathbf{H}_T \mathbf{F}_p^H \mathbf{x} + \mathbf{w}_T, \end{aligned} \tag{7}$$

where $\mathbf{h}_T$, F_p, $\mathbf{x}$, and $\mathbf{w}_T$ are the channel coefficients in TD, the partial Fourier matrix, the OFDM frame in FD and the additive white Gaussian noise (AWGN) in the TD, respectively. Furthermore '$*$' represents the convolution operation, while $\mathbf{H}_T$ is an $(M \times M)$-element circulant matrix generated from the channel vector $\mathbf{h}_T$. It may be noted that (7) is in the TD, whereas (1) is in the FD. In (7), $\mathbf{H}_T\mathbf{F}_p^H$ is an $(M \times N)$ matrix with $M < N$ and hence the equation represents an under-determined system of linear equations, which does not give a unique solution for $\mathbf{x}$. However, since $\mathbf{x}$ is sparse in nature, compressed sensing [29] algorithms can be used for recovering the sparse vector $\mathbf{x}$.

The sufficient condition for the successful reconstruction of a sparse high dimensional vector from an under-determined system of linear equations is that the projection matrix, namely the coefficient matrix of the system, should satisfy the Restricted Isometric Property (RIP) of [29]. In such cases, the sparse vector can be reconstructed at a low computational complexity from an under-determined system of equations. Hence, in order to recover $\mathbf{x}$ from (7), the matrix $\mathbf{H}_T\mathbf{F}_p^H$ should satisfy the Restricted Isometric Property (RIP), where $\mathbf{F}_p$ is the partial Fourier Matrix. It has been shown in [30][Theorem 3.3] that if the partial Fourier matrix (F_p) is constructed from the random choice of the columns of the DFT matrix, then F_p^H satisfies the RIP with a high probability. However, there exist also deterministic approaches capable of generating a partial Fourier matrix satisfying the RIP [31], [32]. Hence, for the present discussion, it is assumed that the appropriate selection of rows was executed so that F_p^H satisfies the RIP. However, the product $\mathbf{H}_T\mathbf{F}_p^H$ may not satisfy the RIP. This problem can be circumvented by exploiting Lemma 1.

Lemma 1: If a random $M \times 1$ vector $\mathbf{h}$ obeys a continuous distribution, then the rank of the circulant matrix $\mathbf{H}$ constructed from $\mathbf{h}$ is M.

Proof: Let $f(z)$ be the associated polynomial of the circulant matrix H. Then $f(z) = h_0 + h_1 z + \ldots + h_{M-1}z^{M-1}$, with the coefficients h_i being the elements of the first column of $\mathbf{H}$ or the elements of the vector $\mathbf{h}$ itself. The rank of an $(M \times M)$ circulant matrix $\mathbf{H}$ is $M - d$, where d is the degree of $gcd\left[f(z), z^M - 1\right]$ [33], where gcd represents the greatest common divisor. Note that there are only two factors for $z^M - 1$, i.e., $f_1(z) = (z-1)$ and $f_2(z) = z^{M-1} + z^{M-2} + \ldots + z + 1$. Therefore, we have to find $gcd\,[f(z), f_1(z)f_2(z)]$, where neither $f_1(z)$ nor $f_2(z)$ are factorizable. Hence, $d > 0$ if and only if $f(z)$ has either of the terms $f_1(z)$ or $f_2(z)$ as its factors. For $f_2(z)$ to be a factor of $f(z)$, this requires that all the elements of $\mathbf{h}$ should be either unity or a constant. Similarly, for $f_1(z)$ to be factor of $f(z)$, the elements of $\mathbf{h}$ should be of the form $\{1\ -1\ 0\ 0\ 0\ \ldots\}$ or $\{1\ -1\ k_1 - k_1\ k_2\ -k_2, \ldots.\}$, where the coefficients k_i are some constants. However, the probability that the elements of $\mathbf{h}$ assume these specific discrete values becomes zero, when it is generated by sampling from a continuous distribution. Hence, it can be concluded that we have $gcd\left[f(z), z^M - 1\right] = 1$ with probability 1. Therefore, the rank of the circulant matrix H will be M with probability 1, when $\mathbf{h}$ follows a continuous distribution. □

Lemma 2: If F_p^H satisfies the RIP of order k, then the product $\mathbf{H}_T F_p^H$ also satisfies the RIP of the same order with probability 1.

Proof: From Lemma 1, it is clear that the circulant matrix $\mathbf{H}_T$, constructed from the channel vector $\mathbf{h}_T$ is a full rank matrix with probability 1. This is because $\mathbf{h}_T$ follows a continuous distribution (Complex Normal). Hence, $\mathbf{H}_T$ is invertible with probability 1. By corollary 3.3 of [34], if a matrix $\mathbf{A}$ satisfies the RIP of order k, then for any invertible matrix $\mathbf{B}$, the product $\mathbf{BA}$ also satisfies the RIP of the same order. Therefore, the product $\mathbf{H}_T F_p^H$ satisfies the RIP of the order k with probability 1. □

From Lemma 2, it is evident that the compressive sensing techniques can be directly applied to recover $\mathbf{x}$ from $\mathbf{y}_T$. Hence, the procedure for the detection of CSMIM is summarized below.

- Apply a compressed sensing algorithm [35]–[37] to recover $\mathbf{x}$ from $\mathbf{y}_T$ using $\mathbf{y}_T = \mathbf{H}_T\mathbf{F}_p^H\mathbf{x} + \mathbf{w}_T$. This reconstruction is possible, since $\mathbf{H}_T\mathbf{F}_p^H$ satisfies the RIP with probability 1.
- Identify the non-zero locations of $\mathbf{x}$. These locations can be used to decode the cluster indices and subcarrier indices in each of the clusters.
- Approximate each non-zero component of $\mathbf{x}$ to the nearest PSK/QAM modulating symbols.

All these three schemes i.e., MIM, DMIM and CSMIM are coherent modulation schemes, which requires accurate CSI for detecting the transmitted signal. The next section describes a non-coherent MIM scheme that does not require the knowledge of the channel to recover the symbols.

IV. NON-COHERENT MULTI LAYER INDEX MODULATION

For non-coherent schemes, no pilot based channel state information (CSI) is used and therefore no pilot overhead is required for estimating the channel. Hence, these schemes are eminently suitable for rapidly fading channels, where

frequent channel estimation would be required, leading to significant pilot overhead. Hence, a non-coherent OFDM-IM scheme was proposed in [23], in which the information is conveyed only through the position of active subcarriers. In this case, a 1 (or equivalently a constant 'θ') is placed in the active subcarriers, but no PSK/QAM symbols are used. Let us assume that there are N subcarriers, which are split into B clusters, each having L subcarriers. In each cluster, K out of L subcarriers are loaded with θs (active subcarriers) and the rest are filled with zeros. The set of active subcarriers is chosen according to the incoming bit sequence. The total number of bits transmitted using this procedure is $B\lfloor \log_2 \left(\binom{L}{K} \right) \rfloor$. The proposed MIM scheme can also be extended to non-coherent detection as follows.

The incoming bit sequence is partitioned into two sets. The first set is used for selecting Q out of B clusters, which are termed as type I clusters. This requires $\lfloor \log_2 \left(\binom{B}{Q} \right) \rfloor$ bits. Within each of the Q Type I clusters, the second bit sequence is used for selecting the active subcarriers, i.e. the subcarriers loaded with θs. In the rest of the $(B-Q)$ clusters, which are termed as Type II clusters, the second bit sequence is used for selecting the passive subcarriers, i.e. the subcarriers which are loaded with zeros. Hence, there will be K non-zero entries in Type I clusters and $(L-K)$ non-zero entries in Type II clusters. The selection of active subcarriers in both Type I and Type II clusters requires $\lfloor \log_2 \left(\binom{L}{K} \right) \rfloor$ bits. Hence, the throughput of this case is $\frac{\lfloor \log_2 \left(\binom{B}{Q} \right) \rfloor + B \lfloor \log_2 \left(\binom{L}{K} \right) \rfloor}{N+N_{CP}}$ bits, whereas it is $\frac{B \lfloor \log_2 \left(\binom{L}{K} \right) \rfloor}{N+N_{CP}}$ in the case of non-coherent Index Modulation. This shows that for non-coherent MIM, there is a throughput advantage over non-coherent IM. However, the energy per bit of a transmission in this case is $E_b = \frac{(KQ+(L-K)(B-Q))E}{\left\lfloor \log_2 \left(\binom{B}{Q} \right) \right\rfloor + B \left\lfloor \log_2 \left(\binom{L}{K} \right) \right\rfloor}$, which will be lower or higher or same as that of the non-coherent IM scheme, depending on the values of K and Q. Observe that E_b can be readily reduced without changing the throughput by choosing $K < L/2$ and $K > L/2$, when $Q > B/2$ and $Q < B/2$, respectively. The basic requirement for MIM to work is that we should have $K \neq \frac{L}{2}$, i.e. the number of active and passive subcarriers should be different. In the scenarios, when they are the same, i.e. $K = L-K$, all blocks are identical, hence it will not be possible to identify Q out of B Type I clusters. This is proved mathematically in Theorem 1. Note that here we used Type I and Type II to distinguish two types of clusters instead of active and passive ones, since all the clusters are active in this case. The detector conceived for the proposed scheme is detailed below.

A. DETECTOR

Mathematically, this system can also be represented using (1) except that in this case $x(n)$ will be either θ or 0. A two-stage ML detector can be employed. In the first stage, the Type I clusters are detected and in the second stage, the activation patterns (subcarriers with non-zero values) in each of the clusters are identified. Let the channel coefficient for all of the subcarriers be independent and identically distributed as $h(n) \sim \mathcal{CN}(0, \sigma_h^2)$. Hence, the distributions of the received data at the n^{th} subcarrier are

$$\begin{aligned} &\Pr(y(n)|x(n)=0) \\ &\quad = \mathcal{CN}(0, \sigma^2), \qquad (8) \\ &\Pr(y(n)|x(n)=\theta) \\ &\quad = \int_{-\infty}^{\infty} \Pr\left(y(n)|x(n)=\theta, h(n)\right) \Pr\left(h(n)\right) dh \\ &\quad = \int_{-\infty}^{\infty} \mathcal{CN}\left(\theta h(n), \sigma^2\right) \mathcal{CN}(0, \sigma_h^2) dh \\ &\quad = \mathcal{CN}(0, \sigma^2 + \theta^2 \sigma_h^2). \qquad (9) \end{aligned}$$

Now, if a particular cluster is Type I, then it will have K θs and $(L-K)$ zeros, while if it is Type II, there will be K zeros and $(L-K)$ θs. Hence, in order to identify whether a particular cluster is Type I or Type II, we have to count the number of θs in that cluster. We now want to differentiate between two cluster, i.e. the cluster with K θs and $(L-K)$ zeros as well as that having K zeros and $(L-K)$ θs. Hence, what we need is the distribution of the number of θs in a cluster given the observations, which is formulated as $\Pr(\mathcal{I}_b|\mathbf{y}_b)$, where $\mathcal{I}_b$ is the number of θs in a cluster $b \in \{0, 2 \ldots, B-1\}$ and $\mathbf{y}_b = \{y(bL), y(bL+1), \ldots, y((b+1)L-1)\}$ represents the received vector in that cluster. This probability can be obtained as follows. Note that it is only the number of θs which is important to us, not the specific positions of θs. Therefore, all possible locations of θs (with the total number of θs being the same) has to be considered. Hence, we have to evaluate

$$\Pr\left(\mathcal{I}_b|\mathbf{y}_b\right) = \Pr\left(\mathbf{y}_b|\mathcal{I}_b\right) \frac{\Pr(\mathcal{I}_b)}{\Pr\left(\mathbf{y}_b\right)}, \qquad (10)$$

where $\mathcal{I}_b = K$ or $\mathcal{I}_b = L-K$. Note that we have $\Pr(\mathcal{I}_b = K) = \frac{Q}{B}$ and $\Pr(\mathcal{I}_b = L-K) = \frac{B-Q}{B}$, since there are Q Type I clusters. Let Ω_K and Ω_{L-K} be the sets of $\mathbf{x}_b$ with exactly K and $(L-K)$ non-zero elements, respectively. Hence, we have

$$\begin{aligned} \Pr\left(\mathbf{y}_b|\mathcal{I}_b=K\right) &= \sum_{\mathbf{s}_i \in \Omega_K} \Pr\left(\mathbf{y}_b|\mathcal{I}_b=K, \mathbf{s}_i\right) \Pr\left(\mathbf{s}_i|\mathcal{I}_b=K\right) \\ &= \sum_{i=1}^{|\Omega_K|} \frac{1}{|\Omega_K|} \frac{1}{v_0} e^{\left(-\sum_{j \in \Gamma_i} \frac{\|y(j)\|^2}{2(\theta^2\sigma_h^2+\sigma^2)} - \sum_{j \in \Gamma_i^c} \frac{\|y(j)\|^2}{2\sigma^2} \right)}, \qquad (11) \end{aligned}$$

where $v_0 = (2\pi)^{L/2} (\sigma^2)^{K/2} (\theta^2\sigma_h^2+\sigma^2)^{\frac{L-K}{2}}$ and Γ_i represents the locations of the non-zero elements of $\mathbf{s}_i$, while Γ_i^c is its complementary set. Note that we have $\Pr\left(\mathbf{s}_i|\mathcal{I}_b=K\right) = \frac{1}{|\Omega_K|}$ for $\mathbf{s}_i \in \Omega_K$, since each element of the set Ω_K is equally likely. Clearly, we have $|\Omega_K| = \binom{L}{K}$, since that many choices are available for exactly K non-zero elements out of L possible locations. It can be seen that the RHS of (11) is in the form of $\sum_{i=1}^{|\Omega_K|} \nu_i \mathcal{CN}(0, \Sigma_i)$, where $\forall\, i$, $\nu_i = \frac{1}{|\Omega_K|}$ and Σ_i is a diagonal matrix with K entries being $\theta^2\sigma_h^2 + \sigma^2$ (i.e., for the locations $\{j\} \in \Gamma_i$), while the remaining of $(L-K)$ entries being σ^2. Hence, $\Pr\left(\mathbf{y}_b|\mathcal{I}_b=K\right)$ is a mixture of multivariate Gaussian distributions (L-dimensional) with all components

having the same weight. We will approximate this mixture distribution by a single multivariate Gaussian one having the same first-order and second-order moments similar to [38] for computational tractability as follows.

The elements in each of the multivariate Gaussian distribution in the RHS of (11) are independent (since Σ_i is diagonal) with different variances, in which $\theta^2\sigma_h^2+\sigma^2$ is the variance of the K elements, i.e. when $j \in \Gamma_i$ and σ^2 is the variance of the rest of the $(L-K)$ elements. Without loss of generality, we let $j=1 \in \Gamma_i$ and then we have $var\left[y(bL)|j=1\in\Gamma_i\right] = \theta^2\sigma_h^2+\sigma^2$. Note that $y(bL)$ is the first element of $\mathbf{y}_b$. Now, given $j=1\in\Gamma_i$, one of the non-zero locations is fixed (i.e., $j=1$) and there are exactly $\binom{L-1}{K-1}$ possible locations for the other $K-1$ non-zero elements, i.e. we have $|\mathbf{s}_i|j=1\in\Gamma_i| = \binom{L-1}{K-1}$. Therefore, in the mixture distribution of (11), there are $\binom{L-1}{K-1}$ components all with $var\left[y(bL)|j=1\in\Gamma_i\right] = \theta^2\sigma_h^2+\sigma^2$. Similarly, when $j=1\in\Gamma_i^c$, we have $var\left[y(bL)\right]=\sigma^2$. In this case, there are $\binom{L-1}{K}$ locations available for the K non-zero locations such that $y(bL)=0$. Hence, in (11), there are $\binom{L-1}{K}$ mixture components with $var\left[y(bL)|j=1\in J_i^c\right]=\sigma^2$. Hence, the unconditional variance of $y(bL)$ is $\frac{\binom{L-1}{K-1}(\theta^2\sigma_h^2+\sigma^2)+\binom{L-1}{K}\sigma^2}{\binom{L}{K}} = \left(\frac{K}{L}\theta^2\sigma_h^2+\sigma^2\right)$. Here we exploited the fact that the variance of a mixture Gaussian distribution is the weighted sum of the variance of each of the component, i.e., $var\left[\sum_i \nu_i\mathcal{N}(0,\sigma_i^2)\right]=\sum_i\nu_i\sigma_i^2$. Note that the other components of $\mathbf{y}_b$, i.e,$j\neq 1$, have the same variances as that of $j=1$, since our interests are not in the specific positions of non-zero elements, but in their number only. Furthermore, the mean of the approximated Gaussian mixture is $\mathbf{0}$, since that of each component in the mixture is $\mathbf{0}$. Hence, now the approximated Gaussian distribution with the same moments as the original distribution is

$$\Pr\left(\mathbf{y}_b|\mathcal{I}_b=K\right)=\mathcal{CN}\left(0,\left(\frac{K}{L}\theta^2\sigma_h^2+\sigma^2\right)\mathbf{I}\right). \quad (12)$$

Similarly, we can show that

$$\Pr\left(\mathbf{y}_b|\mathcal{I}_b=\text{L-K}\right)=\mathcal{CN}\left(0,\left(\frac{L-K}{L}\theta^2\sigma_h^2+\sigma^2\right)\mathbf{I}\right). \quad (13)$$

Hence, from (10), 12) and (13), we have:

$$\Pr\left(\mathcal{I}_b=K|\mathbf{y}_b\right)=\frac{Q}{B}\mathcal{CN}\left(0,\left(\frac{K}{L}\theta^2\sigma_h^2+\sigma^2\right)\mathbf{I}\right) \quad (14)$$

and

$$\Pr\left(\mathcal{I}_b=L-K|\mathbf{y}_b\right) = \frac{B-Q}{B}\mathcal{CN}\left(0,\left(\frac{L-K}{L}\theta^2\sigma_h^2+\sigma^2\right)\mathbf{I}\right). \quad (15)$$

The log likelihood ratio using (14) and (15) is

$$\gamma_b = \frac{\Pr\left(\mathcal{I}_b=K|\mathbf{y}_b\right)}{\Pr\left(\mathcal{I}_b=L-K|\mathbf{y}_b\right)} = \log\left(\left(\frac{(L-K)\beta+L}{L\beta+L}\right)^{\frac{L}{2}}\frac{Q}{B-Q}\right) + \frac{\frac{\beta}{2L\sigma^2}}{\left(\frac{(L-K)}{n}\beta+1\right)\left(\frac{K}{L}\beta+1\right)}(2K-L)\|\mathbf{y}_b\|^2, \quad (16)$$

where $\beta=\frac{\theta^2\sigma_h^2}{\sigma^2}$ is the SNR of the system. Similar to the case of (3), for deriving (16), we assume that $Q<B$. In order to identify whether a cluster is Type I or not, we have to check whether $\gamma_b>0$. However, we know that there are exactly Q Type I clusters. Therefore, rather than testing whether a particular cluster is of Type I, it is proposed to identify Q clusters having the largest γ_b values. The following theorem will formulate our decision rule for cluster identification.

Theorem 1: Recall that the LLR for Type I cluster detection is given by (16). Then, the following holds,

1) If $K>\frac{L}{2}$, then choosing the clusters having the Q largest γ_b values is equivalent to selecting the clusters with the Q highest $\|\mathbf{y}_b\|^2$ values. In other words, $\gamma_{b(1)}\geq\ldots\geq\gamma_{b(Q)} \iff \|y_{b(1)}\|^2\geq\ldots\geq\|y_{b(Q)}\|^2$ where $b(i)\in\{0,1,\ldots,B-1\}$.
2) If $K<\frac{L}{2}$, then choosing the clusters having the Q largest γ_b values is equivalent to selecting clusters having Q lowest $\|\mathbf{y}_b\|^2$ values. In other words, $\gamma_{b(1)}\geq\ldots\geq\gamma_{b(Q)} \iff \|y_{b(1)}\|^2\leq\ldots\leq\|y_{b(Q)}\|^2$.
3) Finally, when we have $K=\frac{L}{2}$, cluster identification fails.

Proof: Consider the first and second terms of the RHS in (16). The first term is a constant and does not affect the maximization operation. All terms of the second term except $(2K-L)$ are positive for all values of K. Now, if $K>L/2$, then $2K-L>0$. Hence in that case, if $\gamma_l>\gamma_j$ for some $l,j\in\{1,\ldots,B\}$, then $\|y_l\|^2>\|y_j\|^2$. This proves the first part. Similarly, if $K<L/2$, then $2K-L<0$. Hence, $\gamma_l>\gamma_j$ implies $\|y_l\|^2<\|y_j\|^2$, which proves the second part of the theorem. Finally, if $K=L/2$, the LLR becomes independent of $\|\mathbf{y}_b\|^2$ and γ_b will be the same for all b and hence the clusters cannot be identified. □

Finally, the detection scheme is summarized as follows.

- First detect the Type I clusters. Depending on the value of K, this can be done by selecting the Q clusters having the largest (or smallest) value of $\|\mathbf{y}_b\|^2$;
- In the Q Type I clusters, the non-zero subcarriers can be identified by selecting the specific K subcarriers with the largest energy (same procedure as in non-coherent index modulation [23]);
- In the remaining $B-Q$ clusters, the Type I subcarriers are identified by choosing the $(L-K)$ subcarriers having the highest energy or equivalently the K subcarriers having the lowest energy;
- Finally, the transmitted bits are decoded from the set of Type I clusters and subcarriers.

B. THEORETICAL ANALYSIS

The probability of error in non-coherent MIM is analysed in this section. An error will occur either because a cluster or a subcarrier is wrongly detected. The probability of error for each of the cases is given below. Throughout the

analysis, we assume that $K > L/2$. The analysis is similar for $K < L/2$.

1) PROBABILITY OF CLUSTER ERROR (P_{ce})

First the probability of cluster error is found. From Theorem 1, we have two cases for cluster identification. For the case $K > L/2$, we select Q clusters having the highest $\|\mathbf{y}_b\|^2$ values. Let Λ be the set of correct Q clusters and Λ^c represent its complementary set. There will be no cluster errors, if

$$\min_{b\in\Lambda} \|\mathbf{y}_b\|^2 > \max_{b\in\Lambda^c} \|\mathbf{y}_b\|^2. \tag{17}$$

Note that $\Lambda \subset \{0, 1, \ldots, B-1\}$ and $\Lambda^c \subset \{0, 1, \ldots, B-1\}$ with $|\Lambda| = Q$ and $|\Lambda^c| = B - Q$, are the set of clusters ($\{b\}$) with $\mathcal{I}_b = K$ and $\mathcal{I}_b = L-K$, respectively. Hence, the distribution of $\mathbf{y}_b$, when $b \in \Lambda$ is $\Pr[\mathbf{y}_b|\mathcal{I}_b = K]$ which is given by (12), while the distribution of $\mathbf{y}_b$ when $b \in \Lambda^c$ is $\Pr[\mathbf{y}_b|\mathcal{I}_b = \text{L-K}]$ which is given by (13). We have to determine the distributions $\Pr\left[\|\mathbf{y}_b\|^2|\mathcal{I}_b = K\right]$ and $\Pr\left[\|\mathbf{y}_b\|^2|\mathcal{I}_b = \text{L-K}\right]$ to estimate the probability of cluster error, which is accomplished as follows.

Note that $\|\mathbf{y}_b\|^2 = \mathbf{y}_b^H\mathbf{y}_b = \|y(bL)\|^2 + \|y(bL+1)\|^2 + \ldots + \|y((b+1)L-1)\|^2$, in which the distribution of each of the terms $y(bL), y(bL+1), \ldots, y((b+1)L-1)$ for $\mathcal{I}_b = K$ is $\mathcal{CN}(0, \frac{k}{L}\theta^2\sigma_h^2 + \sigma^2)$. Hence, the corresponding norm squares, i.e., $\|y(bL)\|^2, \|y(bL+1)\|^2, \ldots, \|y((b+1)L-1)\|^2$ for $\mathcal{I}_b = K$ are distributed as $exp(\lambda_1)$, which is the exponential distribution, where $\frac{1}{\lambda_1} = \frac{K}{L}\theta^2\sigma_h^2 + \sigma^2$ and $\|\mathbf{y}_b\|^2$ is the sum of L identically distributed exponential random variables, which follows the Erlang distribution, i.e. we have $\Pr\left(\|\mathbf{y}_b\|^2|\mathcal{I}_b = K\right) \sim Erlang(L, \lambda_1)$. Similarly, $\Pr\left(\|\mathbf{y}_b\|^2|\mathcal{I}_b = \text{L-K}\right) \sim Erlang(L, \lambda_2)$ with $\frac{1}{\lambda_2} = \frac{L-K}{L}\theta^2\sigma_h^2 + \sigma^2$.

Let $U = \{u_1, u_2, \ldots, u_Q\}$ be the set of $\|y_b\|^2$ for $b \in \Lambda$ and $V = \{v_1, v_2, \ldots, v_{B-Q}\}$ be the set of $\|y_b\|^2$ for $b \in \Lambda^c$. Now there will be no errors, if $u_i > v_j,\ \forall\, i \in \{1, \ldots, Q\}, j \in \{1, \ldots, B-Q\}$. Hence, the probability that there will be no cluster error (P_{cc}) is given by

$$\begin{aligned} P_{\text{cc}} &= \bigcap_{j=1}^{B-Q}\bigcap_{i=1}^{Q} \Pr(v_j < u_i), \\ &= p^{Q(B-Q)}, \end{aligned} \tag{18}$$

where $p = \Pr(v_j < u_i)$ and (18) holds due to the independence of the u_is and v_js. Note that the probability of (17) becoming true is the same as that of (18). Lemma 3 will give the value of p.

Lemma 3: When $u_i \sim Erlang(L, \lambda_1)$ and $v_j \sim Erlang(L, \lambda_1)$, the probability $\Pr(v_j < u_i)$ is

$$p = \mathcal{I}_{\frac{\lambda}{1+\lambda}}(L, L) \tag{19}$$

with the regularized Incomplete Beta function defined as [39]

$$\mathcal{I}_z(u, v) = \frac{\mathcal{B}(z; u, v)}{\mathcal{B}(u, v)} = \frac{1}{\mathcal{B}(u, v)}\int_0^z t^{u-1}(1-t)^{v-1}dt, \tag{20}$$

where $\mathcal{B}(u, v) = \mathcal{B}(1; u, v)$ is the complete Beta function. In (19), we have $\lambda = \frac{\lambda_2}{\lambda_1} = \frac{\frac{K}{L}\beta+1}{\frac{L-K}{L}\beta+1}$, with $\beta = \frac{\theta^2\sigma_h^2}{\sigma^2}$

Proof: See Appendix B for proof. □

Theorem 2: The probability of cluster error (P_{ce}) for non-coherent MIM is

$$P_{\text{ce}} = 1 - \mathcal{I}_{\frac{\lambda}{1+\lambda}}(L, L)^{Q(B-Q)}. \tag{21}$$

Proof: The probability of cluster error is

$$\begin{aligned} P_{\text{ce}} &= 1 - P_{\text{cc}} \\ &= 1 - p^{Q(B-Q)}. \end{aligned} \tag{22}$$

When p in (19) is substituted into (22), we arrive at (21). □

Lemma 4: The probability of cluster error in (21) can be upper bounded as

$$P_{\text{ce}} \le 1 - \left(1 - \frac{\lambda^L}{L\mathcal{B}(L, L)(\lambda-1)(\lambda+1)^{2L-1}}\right)^{Q(B-Q)}. \tag{23}$$

Proof: See Appendix C for proof. □

It may be noted that (21) and (23) are derived for the case of $K > L/2$. It can be shown in a similar way that both the equations remain valid for the case of $K < L/2$, except that in this case $\lambda = \frac{\lambda_1}{\lambda_2}$.

2) PROBABILITY OF INDEX ERROR (P_{ie})

In this section, we will analyse the probability of subcarrier index error. An index error may occur either in a correctly or in a wrongly detected cluster. Firstly, the index error in a correctly detected cluster is found as follows.

In non-coherent MIM, there are two types of clusters. In the first type of clusters, there are K non-zero subcarriers, while in the second type of clusters, the number of non-zero subcarriers is $(L-K)$. It has been shown in [23, (13)] that the probability of index error for the case of non-coherent OFDM-IM modulation is

$$P_{\text{ie}}^{(1)} = 1 - \frac{K}{\sigma_1^2}\int_0^\infty \left(1 - e^{-\frac{x}{\sigma_0^2}}\right)^{L-K} e^{-\frac{Kx}{\sigma_1^2}}dx, \tag{24}$$

where $\sigma_1^2 = \theta^2\sigma_h^2 + \sigma^2$ and $\sigma_0^2 = \sigma^2$ in our case. In [23], (24) has been simplified in terms of a binomial expression. By contrast, we evaluate (24) using the Beta function as follows.

Lemma 5: The probability of index error in (24) is

$$P_{\text{ie}}^{(1)} = 1 - \frac{K}{\beta+1}\mathcal{B}\left(L-K+1, \frac{K}{\beta+1}\right), \tag{25}$$

where $\beta = \frac{\theta^2\sigma_h^2}{\sigma^2}$ and $\mathcal{B}(x, y)$ is the Beta function.

Proof: See Appendix D for proof. □

The same expression can be used in Q Type I clusters. In the remaining $(B-Q)$ Type II clusters, the expression can be

obtained by replacing K in (25) with $L-K$, since the number of non-zero elements is $L-K$, leading to

$$P_{\text{ie}}^{(2)} = 1 - \frac{L-K}{\beta+1}\mathcal{B}\left(K+1, \frac{L-K}{\beta+1}\right). \quad (26)$$

Now, we can find the probability of index error, when a cluster is wrongly selected or missed. Let us assume that $K > L/2$. Hence, if a Type I cluster is erroneously identified as a Type II cluster, i.e. when a Type I cluster is missed, only $(L-K)$ active subcarriers will be identified instead of the total of K active subcarriers. Similarly, when we have $K < L/2$, $(2K-L)$ subcarriers will be additionally identified and hence they will be in error. Therefore the index error probability in both these cases is 1. Hence, the unconditional probability of index error is

$$\begin{aligned} P_{\text{ie}} &= \Pr(\text{Index Error}|\text{Cluster Error})P_{\text{ce}} \\ &\quad + \Pr(\text{Index Error}|\text{No Cluster Error})\,(1-P_{\text{ce}}) \\ &= P_{\text{ce}} + (1-P_{\text{ce}})\left(P_{\text{ie}}^{(1)}\Pr(\text{Type I Cluster})\right. \\ &\quad \left. + P_{\text{ie}}^{(2)}\Pr(\text{Type II Cluster})\right) \\ &= P_{\text{ce}} + (1-P_{\text{ce}})\left(\frac{Q}{B}P_{\text{ie}}^{(1)} + \frac{B-Q}{B}P_{\text{ie}}^{(2)}\right). \end{aligned} \quad (27)$$

C. EFFECT OF PARAMETERS ON THE PERFORMANCE

The pair of parameters to be determined for non-coherent MIM schemes are the number of Type I clusters, i.e. Q and the number of active subcarriers K. Both these parameters have an effect on the throughput and on the error probability, as detailed below.

1) NUMBER OF TYPE I CLUSTERS (Q)

The number of Type I clusters affects the probability of cluster error in (21). Differentiating (21) with respect to Q, we will get,

$$\frac{\partial P_{\text{ce}}}{\partial Q} = -\mathcal{I}_{\frac{\lambda}{1+\lambda}}(L,L)^{Q(B-Q)}\log\left(\mathcal{I}_{\frac{\lambda}{1+\lambda}}(L,L)\right)(B-2Q). \quad (28)$$

Note that in (28), $\log\left(\mathcal{I}_{\frac{\lambda}{1+\lambda}}(L,L)\right) \leq 0$. Hence, the derivative in (28) is positive for $Q < B/2$ and negative otherwise. Therefore, the probability of cluster error is maximum at $Q = B/2$ and it decreases on both side. On the other hand, the contribution of the number of Type I clusters towards the throughput is $\left\lfloor \log_2\binom{B}{Q} \right\rfloor$, which is maximum for $Q = B/2$.

2) NUMBER OF ACTIVE SUBCARRIERS (K)

The value of K affects both the probability of cluster errors and the probability of index errors. It can be deduced from (14) and (15) that when $K = L/2$, the probability distributions of Type I (i.e. cluster with K active subcarriers) and Type II (i.e., clusters with $L-K$ active subcarriers) will become identical. Hence, it becomes impossible to distinguish the Type I and Type II clusters. As a result, the probability of cluster error will be maximum at $K = L/2$. When K moves away from $L/2$ on both side, the two distributions will become separate and therefore, the probability of cluster error will be reduced. This can also be seen by differentiating (21) as,

$$\begin{aligned} \frac{\partial P_{\text{ce}}}{\partial K} &= -Q(B-Q)\mathcal{I}_{\frac{\lambda}{1+\lambda}}(L,L)^{Q(B-Q)-1}\frac{\partial}{\partial K}\mathcal{I}_{\frac{\lambda}{1+\lambda}}(L,L) \\ &= -Q(B-Q)L\mathcal{I}_{\frac{\lambda}{1+\lambda}}(L,L)^{Q(B-Q)-1}\left(\frac{\frac{K}{L}\beta+1}{\beta+1}\right)^{L} \\ &\leq 0. \end{aligned} \quad (29)$$

Since (29) is derived for $K > L/2$, the probability of cluster error decreases, as K increases to L. Similarly, for $K < L/2$, it can be shown that P_{ce} decreases, as K decreases to 0.

Let us now consider the probability of index error. Recall that there are K and $(L-K)$ active subcarriers in Type I and Type II clusters. Hence, if a Type I cluster is identified as a Type II cluster or vice-versa, at least $|(L-2K)|$ subcarriers will be in error. Hence, it can be concluded that index error probability is also maximum at $K = L/2$ and it decreases, when K tends to zero or L. This is evident in Fig. 10 of Section V.

Therefore, it can be concluded that the probability of both types of error will be at its minimum, if K tends to zero or L. However, the number of bits transmitted by subcarrier index selection is $B\left\lfloor \log_2\binom{L}{K} \right\rfloor$, which is maximum for $K = L/2$. Therefore, similar to the case of Q, we have to strike a BER vs. throughput trade-off by appropriately choosing K. Hence, it is recommended to keep the value of K between $L/4$ and $L/2$.

V. SIMULATION RESULTS

First we demonstrate the advantages of our MIM philosophy. Subsequently, the comparative study of our coherent and non-coherent schemes is provided and the simulation results are compared to the analytical results.

A. ADVANTAGES OF THE MULTI LAYER FRAMEWORK

To demonstrate the benefits of the multilayer framework, consider an OFDM-IM system having $N = 512$ subcarriers. These subcarriers can be distributed in many different ways to construct the clusters. Let us consider two such configurations; the first configuration uses $B = 16$ clusters each having $L = 32$ subcarriers and the second one $B = 64$ clusters having $L = 8$ subcarriers. Let us assume that half of the subcarriers are used for transmitting QAM/PSK symbols, i.e. $K = L/2$. The throughputs of an OFDM-IM system having these configurations for $16QAM$ schemes are 2.74 bits/s/Hz and 2.59 bits/s/Hz, respectively. Hence, it can be seen that for better throughput, clusters with more subcarriers are preferred. Let us now consider the following schemes.

1) $S1$ - OFDM-IM with $L = 32$, $B = 16$ and $K = 16$.
2) $S2$ - MIM associated with $L = 32$, $B = 16$, $K = 16$ and $Q = 8$.
3) $S3$ - DMIM with the same configuration as that of $S2$. We used $16QAM$ with 0^0 and 45^0 phase rotations as the two modulations.

TABLE 1. Comparison of various schemes.

Scheme	$S1$	$S2$	$S3$	$S4$	$S5$
TP	2.81	1.43	2.84	1.44	2.77
E_b	0.172	0.169	0.170	0.168	0.175
SNR_2	29	27	29	25	25

4) $S4$ - MIM with two layers. In this scheme, 512 subcarriers are split into 8 clusters each having 64 subcarriers. These 8 clusters as super-clusters. Each of these super-clusters are further divided into 8 clusters, each having 8 subcarriers. In this scheme, the incoming bit sequences are split into four sets. The first set is used to select 4 out of 8 super-clusters and in the resultant 4 super-clusters, we used $16QAM$ with 0^0 phase rotation, while in the other 4 clusters, we used $16QAM$ with 45^0 phase rotation. In each of the super-clusters, MIM associated with $B = 8$, $K = 4$ and $Q = 4$ is used.
5) $S5$ - DMIM version of $S4$. Note that in this scheme, we have used four different modulations schemes, i.e. $16QAM$ with 0^0, 30^0, 60^0 and 90^0 phase rotations.

The above schemes were investigated in a dispersive Rayleigh fading channel having 10 channel taps, which were generated by sampling from $\mathcal{CN}(0, 1/10)$. This represents a dispersive channel, where the energy is distributed across all 10 taps equally. It is assumed that perfect knowledge of the CSI is available at the receiver. The Bit Error Rate (BER) is plotted against E_b/N_0 in Fig. 3. Here, we have $E_b/N_0 = \frac{m_A E}{m\sigma^2}$, where m_A is the actual number of active subcarriers in an OFDM frame, E is the average symbol energy, which is unity, m is the number of bits transmitted in an OFDM frame and σ^2 is the noise variance, which is varied to simulate different SNRs. The BER performance is better for schemes having more layers (i.e. $S4$ and $S5$), compared to the other schemes. TABLE 1 compares the various schemes in terms of TP, E_b and SNR_2, which is the SNR in dB required to achieve a BER close to 0.002. It can be seen from TABLE 1 that there is not much change in the throughput or E_b for the multilayer structure, when compared to the corresponding single layer version. For example, for the DMIM schemes ($S3$ and $S5$), there is a marginal reduction in throughput for $S5$ and consequently a slight increase in the energy per bit compared to $S3$. However, there is more than 4 dB SNR gain for $S5$ over $S3$ at all BERs. The reduction in BER for a multilayer architecture can be explained as follows.

For an OFDM-IM system, the information is embedded also in the selection of message-bearing subcarriers. In this case, the misidentification of a subcarrier contributes to the BER. If we opt for a multilayer architecture, additional information is mapped to the selection of clusters. The probability of cluster error is lower than the index error probability, since more than one subcarrier indices have to be selected erroneously for a cluster to be in error. As the number of layers increases, the information carried by the cluster selection increases and the number of information-bearing subcarrier

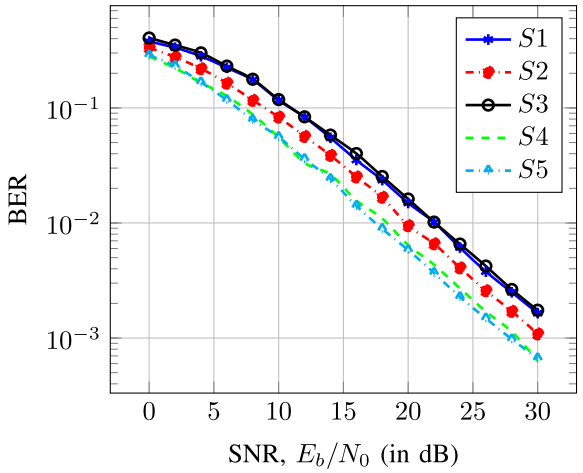

FIGURE 3. BER vs. E_b/N_0 performance of systems $S1$, $S2$, $S3$, $S4$ and $S5$.

indices reduces. For example, in the current simulation set up, there are 32 subcarriers in a cluster for schemes $S2$ and $S3$, which reduces to 8 when we introduce an additional layer (schemes $S4$ and $S5$). Hence, the number of possible index selection combinations reduces from $\binom{32}{16} \approx 6 \times 10^8$ to $\binom{8}{4} = 70$ in the case of $S4$ and $S5$. This also reduces the probability of index error. Hence, the overall BER performance is improved, when the multilayer framework is introduced. On the other hand, if the number of subcarriers in a cluster is reduced in OFDM-IM, the throughput is reduced.

A second advantage of the multi layer framework is the reduction of complexity of encoding and decoding of index modulation. For example, for the simulation problem, for a cluster selecting 16 out of 32 active subcarriers, we have to select a subcarrier activation pattern from $\binom{32}{16} \approx 6 \times 10^8$ possible combinations. In this case, using a look up table based method becomes infeasible, and we have to opt for the combinatoric approach of [6]. On the other hand, there are only 64 combinations for schemes $S4$ and $S5$, which can be readily implemented using the look up table approach. However, it may be noted that when the number of layers increases, we have to carry out the detection in each of the additional layer, which calls for computing the LLR values in those layers. This will result in a modest increase in the computational complexity.

Therefore, in our multilayer framework, we have 1) Better BER performance without sacrificing the throughput and 2) Reduced encoding and decoding complexity. The next subsections provide the performance comparisons of various schemes. We restrict our simulations of the multilayer framework to a single additional layer.

B. PERFORMANCE EVALUATION OF COHERENT SYSTEMS

An OFDM-IM system with the parameters in TABLE 2 is considered for our simulations. We have used $16QAM$ for (MIM and CSMIM) and a combination of $16QAM$ with 0^0 and 45^0 phase rotated scheme for DMIM.

TABLE 2. Simulation parameters.

Parameter	Value
FFT length	128
Cyclic prefix length N_{CP}	16
No of clusters B	8
No of subcarriers in a cluster L	16
No of active subcarriers K	4
No of active clusters Q	4

TABLE 3. Comparison of various coherent schemes (with perfect CSI).

Scheme	OFDM	IM	MIM	DMIM	CSMIM
TP	1.78	1.44	0.76	1.49	1.38
E_b	0.250	0.154	0.146	0.149	0.075
SNR_2	24	27	25	30	29

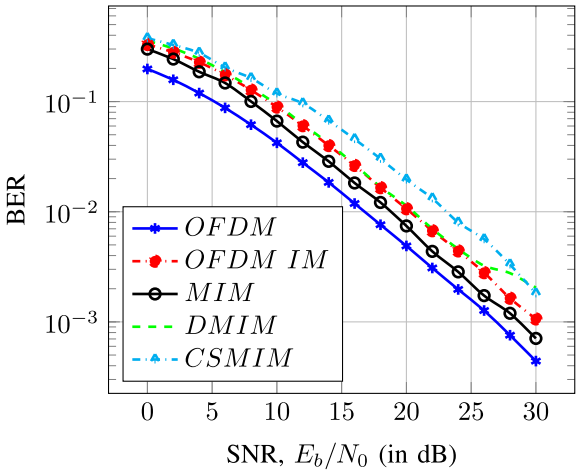

FIGURE 4. BER vs. E_b/N_0 - Coherent systems (with perfect CSI).

The channel used is the same as that of Section V-A. It is assumed that perfect knowledge of the CSI is available at the receiver. The results are shown for 1) OFDM system having similar data rate (with 32 subcarriers), 2) OFDM-IM, 3) MIM 4) DMIM and 5) CSMIM with half the number of TD samples (i.e 64 TD samples). We have also simulated the performance in the face of channel estimation errors. This was emulated by adding noise having a variance of σ^2 (variance of AWGN noise) to the true channel coefficients. This is because typically channel estimation is carried out using the classic least square approach and this leads to channel estimates that are near-Gaussian distributed with $\hat{h}_T \sim \mathcal{CN}(h_T, \sigma^2_{h_T})$, where $\sigma^2_{h_T}$ depends on the variance of the underlying noise in the system and on the number of observations available for channel estimation [40].

The BER is plotted against E_b/N_0 in Fig. 4 for the scenario, when perfect CSI is available, while TABLE 3 shows the comparison of various coherent schemes in terms of TP, E_b and SNR_2. OFDM gives the best performance in terms of both BER and throughput. The BER of MIM is better than OFDM-IM, while that of DMIM and CSMIM is high as expected. The throughput of DMIM is higher than that of OFDM-IM and lower than than that of OFDM, while that of CSMIM is lower than that of OFDM-IM. However, when

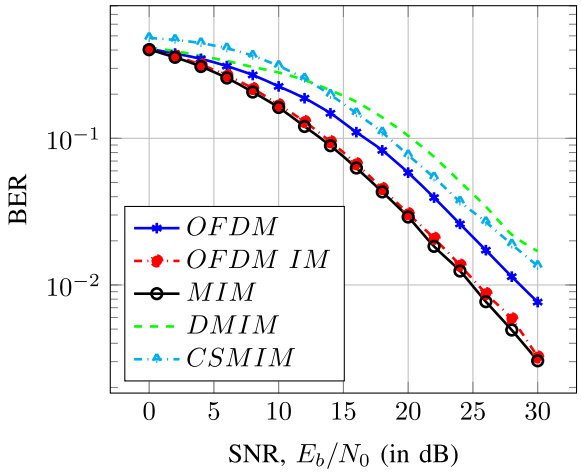

FIGURE 5. BER vs. E_b/N_0 - Coherent systems (with CSI error).

E_b is compared, it is best for CSMIM and worse for OFDM. This means that the best BER of OFDM is attained at the cost of its high energy requirement. The poor performance of CSMIM is due to the active subcarriers sparsity of our example. The number of non-zero subcarriers in the system is 16 out of 128, which have to be recovered from 64 TD samples. All the three proposed schemes outperform both OFDM and OFDM-IM in terms of their energy efficiency. Fig. 5 shows the BER performance of these schemes in the face of CSI errors. It can be seen that in the presence of CSI errors, the performance of OFDM is degraded. In this case, OFDM-IM and MIM give the best performance. This is because, in conventional OFDM, the information is only contained in the classic modulated symbols (e.g. QAM16), which have higher probability of getting corrupted in the face of CSI errors, since there are 16 legitimate symbols. By contrast, in OFDM-IM and MIM, part of the information bits are embedded in the selection of sub-carriers, where there are only two options (either the sub-carrier is loaded with a classic symbol or not) and hence there is a lower probability of erroneous decisions, for example owing to CSI errors.

C. PERFORMANCE EVALUATION OF NON-COHERENT SYSTEMS

The performance of our non-coherent IM system has been studied in the following three channel models.

- $C1$ - A highly frequency selective channel, in which the FD channel coefficients are uncorrelated and have unit variance, i.e. $h \sim (0, \mathbf{I})$. This channel is selected for comparing our theoretical analysis to the simulation results.
- $C2$ - 10-tap dispersive Rayleigh fading channel, where the taps were generated by sampling from $\mathcal{CN}(0, 1/10)$. This will make the variance of the channel coefficient in FD unity, i.e. $\sigma^2_h = 1$. This channel is exactly the same as that used in our coherent system simulations.
- $C3$ - A flat fading channel.

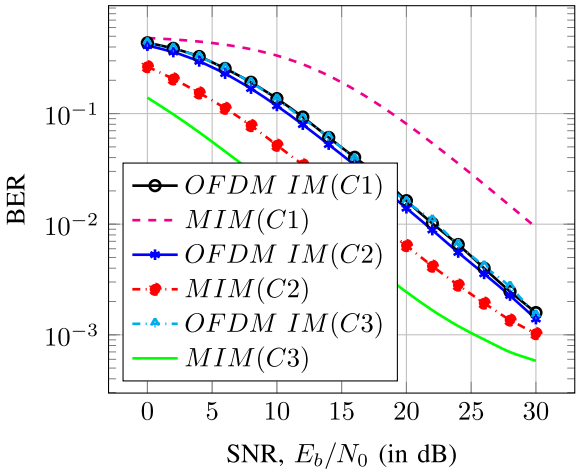

FIGURE 6. BER vs. E_b/N_0 - Non-coherent systems.

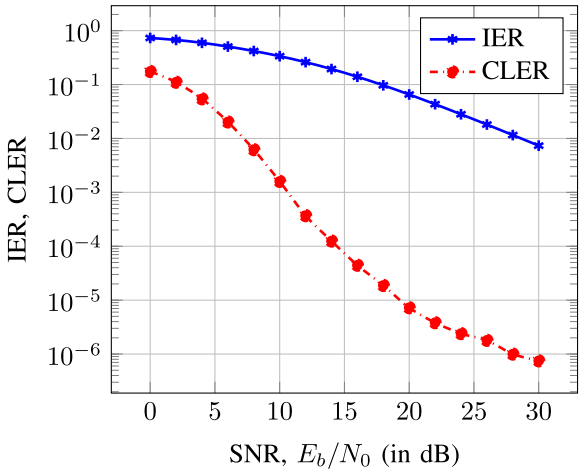

FIGURE 7. IER/CLER vs. E_b/N_0.

TABLE 4. Comparison of the non-coherent schemes.

Scheme	IM-NC	MIM-NC
TP	0.50	0.53
E_b	0.33	0.58
SNR_2	27	25

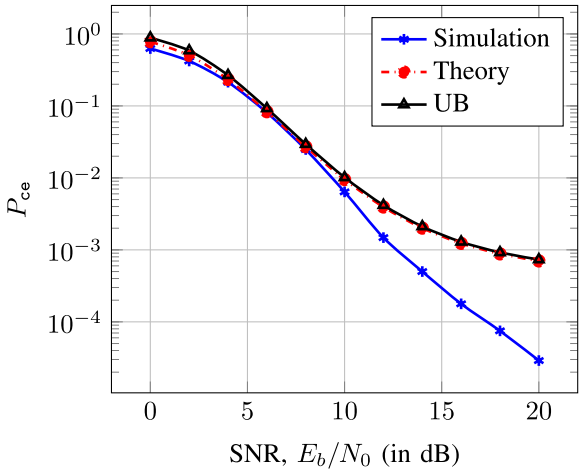

FIGURE 8. P_{ce} vs. E_b/N_0. The upper bound is based on (23), while the theory on (21).

Similar to coherent system simulations, we used an OFDM system of 128 subcarriers, which are split into 8 clusters each having 16 subcarriers. In this case, no PSK/QAM scheme is used and no CSI is required for detection. We kept $K = 3$, $Q = 2$ and $\theta = 1$ for our simulations. Fig. 6 shows the BER plotted against E_b/N_0 for both the channel models, where $E_b = \frac{(KQ+(L-K)(B-Q))\sigma_h^2}{\lfloor\log_2\left(\binom{B}{Q}\right)\rfloor+B\lfloor\log_2\left(\binom{L}{K}\right)\rfloor}$ is the energy per bit. The comparison of the two non-coherent schemes in terms of TP, E_b and SNR_2 (for channel C_2) is given in TABLE 4. It can be seen that the proposed MIM-NC scheme provides better throughput and BER than IM-NC, albeit at the cost of a high E_b.

The performance of our proposed scheme improves substantially when the channel changes from a highly frequency selective channel ($C1$) to a flat fading channel ($C3$). This is because in MIM part of the message bits are embedded into clusters by dividing them into two groups, viz. Type I and Type II. The cluster error probability (P_{ce}) increases, when the channel changes from flat to frequency selective. The increase in P_{ce} can be explained as follows. In an OFDM frame, the Type I clusters are identified by selecting Q clusters having either maximum or minimum energy (Refer Theorem 1) and this maximum or minimum remains the same for a flat fading channel, since the FD channel coefficients ($h(n)$) remain the same across all sub-carriers. When the channel starts behaving as frequency selective, $h(n)$ starts changing slowly and this can affect the maximum or minimum. Hence, for a channel like $C1$, there is a high probability that the clusters having maximum or minimum energy are wrongly identified. The error in cluster identification will cause the message bits embedded in the clusters and in the sub-carriers of the wrongly selected clusters become erroneous and will result in high BER. For OFDM IM, there are no bits embedded in the clusters and its performance is almost the same across all channels. Therefore, the advantage of MIM over OFDM IM will decrease, as the channel variation increases.

Fig. 7 shows the Index Error Rate (IER) and Cluster Error Rate (CLER), which are defined as the number of indices in error normalized by the total number of indices and the number of clusters in error to the total number of clusters, respectively. It can be seen that the CLER is much lower than the IER. This essentially brings about the advantage of the multi layer architecture detailed in Section V-A. Explicitly, if the bits are mapped in a cluster, rather than to a subcarrier, the BER will be improved.

Fig. 8 shows the cluster error probability, i.e. probability that there is at least a cluster is in error in a OFDM frame. Note that this is different from the CLER. Fig. 8 shows P_{ce} obtained through simulations, theoretical analysis (21) and

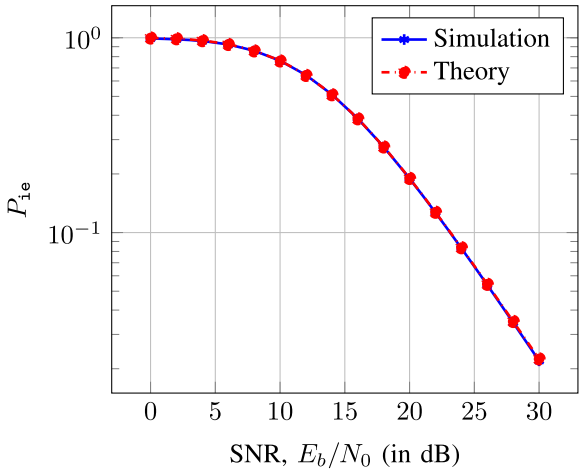

FIGURE 9. P_{ie} vs. E_b/N_0.

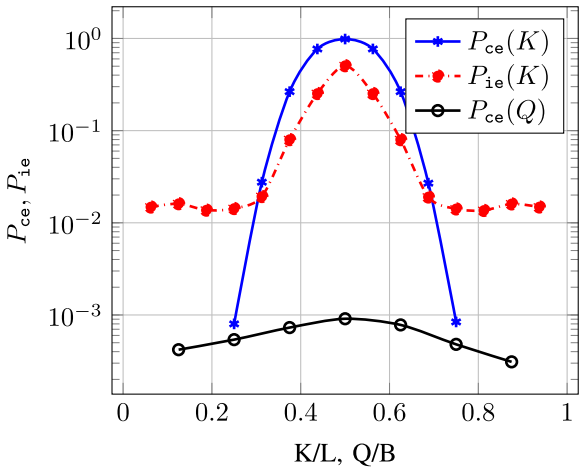

FIGURE 10. Effect of the number of active subcarriers and clusters.

its upper bound (23). The figure shows close match between the simulation results and the theoretical values of P_{ce} at low SNRs. At high SNRs, a gap can be seen between the simulation results and the theoretical values. However, it should be noted that at these SNR values, the probability of cluster error is lower than 10^{-3}, where the disparity is due to the approximation of a Gaussian mixture by a single Gaussian distribution. However, it should be noted that the upper bound derived for P_{ce} is tight. Note that the theoretical cluster probability error flattens out as the SNR increases. This is because the parameter $\lambda = \frac{\frac{K}{L}\beta+1}{\frac{L-K}{L}\beta+1}$ in the expression of P_{ce} (21) approaches $\frac{K}{L-K}$, when the SNR increases. Similarly, Fig. 9 shows the probability of index error (P_{ie}), i.e. the probability that at least one subcarrier index is in error in an OFDM frame. In this case, the theoretical value closely follows the simulation results.

Fig. 10 shows effect of parameters on the performance, which is discussed in Section IV-C. Explicitly, the figure shows the cluster error probability and index error probability vs. the number of active subcarriers (K) and the number of Type I clusters (Q). It can be seen that P_{ce} reaches its maximum for $K = L/2$ and hence naturally decreases on both sides. Similarly, P_{ie} is also maximum for the same value of K. Fig. 10 also shows the dependence of P_{ce} on Q, where P_{ce} is maximum at $Q = B/2$ and it decreases on both sides. Since P_{ie} does not directly depend on Q, it is not shown. The results of Fig. 10 are congruent with the discussions of Section IV-C.

It can be seen from TABLE 3 that MIM offers better E_b and BER performance than OFDM-IM, however at a lower throughput. By contrast, DMIM has a better throughput and E_b. CSMIM is highly energy efficient and its performance will be improved if the number of active subcarriers is low, which will effectively reduce the throughput. Finally, observe from TABLE 1 that if the number of layers is increased, the BER performance is improved, but there will be a marginal reduction in the throughput and E_b. Hence, it can be concluded from TABLE 1, 3 and 4 that the proposed multilayer architecture is capable of flexibly controlling the throughput, E_b and BER.

VI. CONCLUSION

We proposed a multilayer framework of coherent OFDM-IM to convert a long OFDM frame into multilevel clusters. This multilayer index modulation provides an additional set of parameters, i.e. the number of active clusters in each layer, which can be controlled to strike a trade off between the BER performance, throughput and energy efficiency. As the number of layers increases, more bits are embedded in the clusters and hence we achieve a lower BER, albeit at the cost of some minor reduction in the throughput and energy efficiency. The multilayer framework is generic in nature, hence it may be applied to many of the IM schemes available in the literature.

We also proposed two variants for the MIM, viz. DMIM and CSMIM to improve the throughput of coherent IM. We established the advantages of the proposed architecture through simulations. The idea has also been successfully extended to the case of non-coherent IM, which provides an improved BER and throughput, although at the cost of high E_b. The advantage of MIM-NC over non-coherent IM schemes increases, when the frequency selectivity of the channel reduces. The detectors of the proposed schemes have also been derived and studied by simulations. Furthermore, the probability of error was quantified for our non-coherent system and validated by simulations.

APPENDIX A
DERIVATION OF THE LLR FOR ACTIVE CLUSTER DETECTION

Let $\mathcal{M}$ be the set of PSK/QAM constellation symbols and $\mathbf{y}_b$ denote the received data vector in the b^{th} cluster, i.e., $\mathbf{y}_b = [y(bL), y(bL+1), \ldots, y((b+1)L-1)]^T$. Let $b = 1$ and $b = 0$ indicate whether the b^{th} cluster is active or not, respectively.

The likelihood ratio is

$$\gamma_b = \log\left(\frac{\Pr(b=1|\mathbf{y}_b)}{\Pr(b=0|\mathbf{y}_b)}\right) = \log\left(\frac{\prod_{n=bL}^{(b+1)L-1}\Pr(y(n)|b=1)}{\prod_{n=bL}^{(b+1)L-1}\Pr(y(n)|b=0)}\frac{\Pr(b=1)}{\Pr(b=0)}\right) \quad (30)$$

We can assume that all the clusters are equally likely to be active, since no prior knowledge is available. Hence, we have $\Pr(b=1)=\frac{Q}{B}$ and $\Pr(b=0)=\frac{B-Q}{B}$. Let Γ be the set of active subcarriers in a cluster. We have $\Pr(y(n)|b=0)=\mathcal{CN}(0,\sigma^2)$ and

$$\begin{aligned}\Pr(y(n)|b=1) &= \sum_{s_i\in\mathcal{M}}\Pr(y(n)|b=1,x(n)=s_i)\Pr(x(n)=s_i|b=1)\\ &\quad+\Pr(y(n)|b=1,x(n)=0)\Pr(x(n)=0|b=1), \quad (31)\end{aligned}$$

where $\Pr(x(n)=s_i|b=1)=\Pr(x(n)=s_i|b=1,n\in\Gamma)\Pr(n\in\Gamma|b=1)=\frac{1}{M}\frac{K}{L}$ and $\Pr(x(n)=0|b=1)=\frac{L-K}{L}$ (Note that there are K subcarriers conveying PSK/QAM symbols). Hence, we arrive at:

$$\Pr(y(n)|b=1)=\frac{K}{ML}\sum_{s_i\in\mathcal{M}}\mathcal{CN}(s_ih(n),\sigma^2)+\frac{L-K}{L}\mathcal{CN}(0,\sigma^2). \quad (32)$$

Finally, (30) becomes

$$\begin{aligned}\gamma_b &= \log\left(\frac{Q}{B-Q}\right)+\sum_{n=bL}^{(b+1)L-1}\log\left(\frac{\Pr(y(n)|b=1)}{\Pr(y(n)|b=0)}\right)\\ &= \log\left(\frac{Q}{B-Q}\right)+\sum_{n=bL}^{(b+1)L-1}\log(\eta(n))\\ &= \log\left(\frac{Q}{B-Q}\right)+\sum_{n=bL}^{(b+1)L-1}\log\\ &\quad\times\left(\frac{L-K}{L}+\frac{K}{ML}\sum_{s_i\in\mathcal{M}}e^{-\frac{\left(\|y(n)-s_ih(n)\|^2-\|y(n)\|^2\right)}{2\sigma^2}}\right). \quad (33)\end{aligned}$$

where

$$\eta(n)=\frac{\frac{K}{ML}\sum_{s_i\in\mathcal{M}}e^{-\frac{\|y(n)-s_ih(n)\|^2}{2\sigma^2}}+\frac{L-k}{L}e^{-\frac{\|y(n)\|^2}{2\sigma^2}}}{e^{-\frac{\|y(n)\|^2}{2\sigma^2}}}$$

Rearranging (33) will give (3).

APPENDIX B
PROOF OF LEMMA 3

Let $f_u(u)$ be the density function of $u\in U$ and $F_v(v)$ be the distribution function of $v\in V$. Now, we can estimate p as

$$\begin{aligned}p &= \Pr(v_j<u_i)\\ &= \int_0^\infty \Pr(v_j<u_i|u_i=u)f_{u_i}(u)du\\ &= \int_0^\infty F_v(u)f_U(u)du\\ &= \int_0^\infty\left(1-\sum_{l=0}^{L-1}\frac{1}{l!}\mathbf{e}^{-\lambda_2 u}(\lambda_2 u)^l\right)f_u(u)du\\ &= 1-\int_0^\infty\sum_{l=0}^{L-1}\frac{1}{l!}\mathbf{e}^{-\lambda_2 u}(\lambda_2 u)^l f_u(u)du\\ &= 1-\int_0^\infty\sum_{l=0}^{L-1}\frac{1}{l!}\mathbf{e}^{-\lambda_2 u}(\lambda_2 u)^l\frac{\lambda_1^L}{(L-1)!}e^{-\lambda_1 u}u^{L-1}du\\ &\overset{(a)}{=} 1-\sum_{l=0}^{L-1}\frac{\lambda_1^L\lambda_2^l}{l!(L-1)!}\int_0^\infty\mathbf{e}^{-(\lambda_1+\lambda_2)u}u^{l+L-1}du\\ &\overset{(b)}{=} 1-\sum_{j=0}^{L-1}\frac{\lambda_1^L\lambda_2^l}{l!(L-1)!}\frac{(l+L-1)!}{(\lambda_1+\lambda_2)^{l+L}}\\ &= 1-\sum_{l=0}^{L-1}\frac{(l+L-1)!}{l!(L-1)!}\frac{\lambda_1^L\lambda_2^l}{\lambda_1^{l+L}\left(1+\frac{\lambda_2}{\lambda_1}\right)^{l+L}}\\ &= 1-\sum_{l=0}^{L-1}\frac{(l+L-1)!}{l!(L-1)!}\frac{\lambda^l}{(1+\lambda)^{l+L}},\\ &= 1-\sum_{l=0}^{n-1}\frac{(l+L-1)!}{l!(L-1)!}\left(\frac{\lambda}{1+\lambda}\right)^l\left(\frac{1}{1+\lambda}\right)^L, \quad (34)\end{aligned}$$

where $\lambda=\frac{\lambda_2}{\lambda_1}$. In (34), (a) is arrived at by changing the order of integration and summation, while (b) is due to the identity $\int_0^\infty e^{-ax}x^m dx=\frac{m!}{a^{m+1}}$. Now, note that the term inside the sum in the RHS of (34) is the probability mass function of a random variable obeying the negative binomial distribution. Explicitly $\left(\frac{\lambda}{1+\lambda}\right)^l\left(\frac{1}{1+\lambda}\right)^L=\Pr(X=l)$, where X follows the negative binomial distribution associated with the parameters $\frac{\lambda}{1+\lambda}$ and L. Therefore the sum is the cumulative mass function of X. Hence, we can express p as

$$\begin{aligned}p &= 1-\Pr(X\le L-1)\\ &= 1-\left(1-\mathcal{I}_{\frac{\lambda}{1+\lambda}}(L,L)\right), \quad (35)\end{aligned}$$

where $\left(1-\mathcal{I}_{\frac{\lambda}{1+\lambda}}(L,L)\right)$ is the cumulative mass function of X upto $L-1$ terms. Lemma 3 directly follows from (35).

APPENDIX C
PROOF OF LEMMA 4

Consider the expression for p in (19). Note that in (19), we have $0.5\le\frac{\lambda}{1+\lambda}\le 1$, since $\lambda\ge 1$. Therefore, from Lemma 6 of Appendix E, a lower bound for p is,

$$\begin{aligned}p &\ge 1-\frac{\left(\frac{\lambda}{(1+\lambda)^2}\right)^L}{\mathcal{B}(L,L)\left(\frac{\lambda-1}{\lambda+1}\right)L}\\ &= 1-\frac{\lambda^L}{L\mathcal{B}(L,L)(\lambda-1)(\lambda+1)^{2L-1}}. \quad (36)\end{aligned}$$

Note that the bound is always true, since $\frac{\lambda}{1+\lambda} \geq \frac{L}{2L} = 0.5$. Finally, substituting the lower bound for p from (36) into (22) will give the upper bound for the probability of cluster error (23), which proves the Lemma.

APPENDIX D
PROOF OF LEMMA 5

We let $z = 1 - e^{-\frac{x}{\sigma_0^2}}$. Then we have,

$$e^{-\frac{Kx}{\sigma_1^2}} = \left(e^{-\frac{x}{\sigma_0^2}}\right)^{\frac{K\sigma_0^2}{\sigma_1^2}} = (1-z)^{K\frac{\sigma_0^2}{\sigma_1^2}}. \tag{37}$$

Now, let us apply a change of variables $z = 1 - e^{-\frac{x}{\sigma_0^2}}$ in the integral of (24) so that $dz = \frac{1}{\sigma_0^2} e^{-\frac{x}{\sigma_0^2}} dx$. Now we can express the probability of index error as

$$\begin{aligned} P_{\text{ie}}^{(1)} &= 1 - K\frac{\sigma_0^2}{\sigma_1^2}\int_0^1 z^{L-K}(1-z)^{K\frac{\sigma_0^2}{\sigma_1^2}-1} dz, \\ &= 1 - \frac{K}{\beta+1}\int_0^1 z^{L-K}(1-z)^{\frac{K}{\beta+1}-1} dz, \end{aligned} \tag{38}$$

where $\frac{\sigma_0^2}{\sigma_1^2} = \frac{1}{\frac{\theta^2\sigma_h^2}{\sigma^2}+1}$. Finally we have, $\int_0^1 z^{L-K}(1-z)^{\frac{K}{\beta+1}-1} dz = \mathcal{B}\left(L-K+1, \frac{K}{\beta+1}\right)$ is the Beta function. This proves the Lemma.

APPENDIX E
BOUND OF INCOMPLETE BETA FUNCTION

Lemma 6: The incomplete Beta function $\mathcal{I}_z(a,b)$ is lower bounded as

$$\mathcal{I}_z(a,b) \geq 1 - \frac{z^a(1-z)^b}{\mathcal{B}(a,b)\,(a-(a+b)(1-z))},\ z > \frac{b}{a+b} \tag{39}$$

Proof: We have the following integral representation of the incomplete Beta function.

$$\begin{aligned} \mathcal{I}_z(a,b) &= \frac{1}{\mathcal{B}(a,b)}\int_{t=0}^{z} t^{a-1}(1-t)^{b-1}dt \\ &= \frac{1}{\mathcal{B}(a,b)}\left(\int_{t=0}^{1} t^{a-1}(1-t)^{b-1}dt \right. \\ &\quad \left. - \int_{t=z}^{1} t^{a-1}(1-t)^{b-1}dt\right) \\ &= \frac{1}{\mathcal{B}(a,b)}\left(\mathcal{B}(a,b) - \int_{t=z}^{1} t^{a-1}(1-t)^{b-1}dt\right) \\ &= 1 - \frac{1}{\mathcal{B}(a,b)}\int_{t=z}^{1} t^{a-1}(1-t)^{b-1}dt \\ &= 1 - \frac{1}{\mathcal{B}(a,b)}\int_{t=0}^{1-z} (1-u)^{a-1}u^{b-1}dt \\ &\overset{(a)}{=} 1 - \frac{1}{\mathcal{B}(b,a)}\int_{t=0}^{1-z} (1-u)^{a-1}u^{b-1}dt \\ &= 1 - \mathcal{I}_{1-z}(b,a), \end{aligned} \tag{40}$$

where (a) is due to the symmetric property of beta function. The last equality is due to the change of variable $u = 1 - t$. Now, we have the following upper bound for the incomplete Beta function [41, (27)]

$$\mathcal{I}_y(a,b) \leq \frac{y^a(1-y)^b}{\mathcal{B}(a,b)\,(a-(a+b)y)},\ y < \frac{a}{a+b}. \tag{41}$$

Applying (41) to $\mathcal{I}_{1-z}(b,a)$, we have,

$$\mathcal{I}_{1-z}(b,a) \leq \frac{z^a(1-z)^b}{\mathcal{B}(a,b)\,(a-(a+b)(1-z))},\ z > \frac{b}{a+b}. \tag{42}$$

Finally, substituting (42) into (40) proves the Lemma. □

REFERENCES

[1] E. Ba ar, "Index modulation techniques for 5G wireless networks," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 168–175, Jul. 2016.
[2] E. Basar, M. Wen, R. Mesleh, M. Di Renzo, Y. Xiao, and H. Haas, "Index modulation techniques for next-generation wireless networks," *IEEE Access*, vol. 5, pp. 16693–16746, 2017.
[3] N. Ishikawa, S. Sugiura, and L. Hanzo, "Subcarrier-index modulation aided OFDM—Will it work?" *IEEE Access*, vol. 4, pp. 2580–2593, 2016.
[4] S. Gokceli, E. Basar, M. Wen, and G. K. Kurt, "Practical implementation of index modulation-based waveforms," *IEEE Access*, vol. 5, pp. 25463–25473, 2017.
[5] P. K. Frenger and N. A. B. Svensson, "Parallel combinatory OFDM signaling," *IEEE Trans. Commun.*, vol. 47, no. 4, pp. 558–567, Apr. 1999.
[6] E. Ba ar, U. Aygölü, E. Panayırcı, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5536–5549, Nov. 2013.
[7] T. Mao, Z. Wang, Q. Wang, S. Chen, and L. Hanzo, "Dual-mode index modulation aided OFDM," *IEEE Access*, vol. 5, pp. 50–60, 2017.
[8] B. Zheng, F. Chen, M. Wen, F. Ji, H. Yu, and Y. Liu, "Low-complexity ML detector and performance analysis for OFDM with in-phase/quadrature index modulation," *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 1893–1896, Nov. 2015.
[9] Q. Li, M. Wen, H. V. Poor, and F. Chen, "Information guided precoding for OFDM," *IEEE Access*, vol. 5, pp. 19644–19656, 2017.
[10] M. I. Kadir, H. Zhang, S. Chen, and L. Hanzo, "Entropy coding aided adaptive subcarrier-index modulated OFDM," *IEEE Access*, vol. 6, pp. 7739–7752, 2018.
[11] M. Wen, B. Ye, E. Basar, Q. Li, and F. Ji, "Enhanced orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4786–4801, Jul. 2017.
[12] E. Ba ar, "OFDM with index modulation using coordinate interleaving," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 381–384, Aug. 2015.
[13] J. Choi and Y. Ko, "TCM for OFDM-IM," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 50–53, Feb. 2018.
[14] Y. Liu, F. Ji, M. Wen, D. Wan, and B. Zheng, "Vector OFDM with index modulation," *IEEE Access*, vol. 5, pp. 20135–20144, 2017.
[15] M. Chafii, J. P. Coon, and D. A. Hedges, "DCT-OFDM with index modulation," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1489–1492, Jul. 2017.
[16] J. Zhang, M. Zhao, J. Zhong, P. Xiao, and T. Yu, "Optimised index modulation for filter bank multicarrier system," *IET Commun.*, vol. 11, no. 4, pp. 459–467, 2017.
[17] E. Ozturk, E. Basar, and H. A. Cirpan, "Generalized frequency division multiplexing with index modulation," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2016, pp. 1–6.
[18] T. Mao, Q. Wang, J. Quan, and Z. Wang, "Zero-padded orthogonal frequency division multiplexing with index modulation using multiple constellation alphabets," *IEEE Access*, vol. 5, pp. 21168–21178, 2017.
[19] T. Datta, H. S. Eshwaraiah, and A. Chockalingam, "Generalized space-and-frequency index modulation," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 4911–4924, Jul. 2016.
[20] E. Ba ar, "Multiple-input multiple-output OFDM with index modulation," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2259–2263, Dec. 2015.
[21] E. Basar, "On multiple-input multiple-output OFDM with index modulation for next generation wireless networks," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 3868–3878, Aug. 2016.

[22] B. Zheng, M. Wen, E. Basar, and F. Chen, "Multiple-input multiple-output OFDM with index modulation: Low-complexity detector design," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2758–2772, Jun. 2017.
[23] J. Choi, "Noncoherent OFDM-IM and its performance analysis," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 352–360, Jan. 2018.
[24] T. Mao, Q. Wang, Z. Wang, and S. Chen, "Novel index modulation techniques: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 315–348, 1st Quart., 2018.
[25] X. Zhang, Q. Wang, R. Zhang, S. Chen, and L. Hanzo, "Performance analysis of layered ACO-OFDM," *IEEE Access*, vol. 5, pp. 18366–18381, 2017.
[26] X. Zhang, Z. Babar, R. Zhang, S. Chen, and L. Hanzo, "Multi-class coded layered asymmetrically clipped optical OFDM," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 578–589, Sep. 2018.
[27] R. Zhang and L. Hanzo, "Multi-layer modulation for intensity-modulated direct-detection optical OFDM," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 5, no. 12, pp. 1402–1412, Dec. 2013.
[28] H. Zhang, L.-L. Yang, and L. Hanzo, "Compressed sensing improves the performance of subcarrier index-modulation-assisted OFDM," *IEEE Access*, vol. 4, pp. 7859–7873, Oct. 2016.
[29] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[30] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure Appl. Math.*, vol. 61, no. 8, pp. 1025–1045, 2008.
[31] P. Xia, S. Zhou, and G. B. Giannakis, "Achieving the Welch bound with difference sets," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1900–1907, May 2005.
[32] G. Xu and Z. Xu, "Compressed sensing matrices from Fourier matrices," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 469–478, Jan. 2015.
[33] A. W. Ingleton, "The rank of circulant matrices," *J. London Math. Soc.*, vol. 1, no. 4, pp. 445–460, 1956.
[34] J. Yi and G. Tan, "Invariance of the spark, NSP order and RIP order under elementary transformations of matrices," 2015, *arXiv:1502.02874*. [Online]. Available: https://arxiv.org/abs/1502.02874
[35] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Conf. Rec. 27th Asilomar Conf. Signals, Syst. Comput.*, vol. 1, Nov. 1993, pp. 40–44.
[36] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.
[37] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.
[38] U. Orguner and M. Demırekler, "Analysis of single Gaussian approximation of Gaussian mixtures in Bayesian filtering applied to mixed multiple-model estimation," *Int. J. Control*, vol. 80, no. 6, pp. 952–967, 2007. doi: 10.1080/00207170701261952.
[39] G. E. Andrews, R. Askey, and R. Roy, *Special Functions*, vol. 71. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[40] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *Proc. IEEE 45th Veh. Technol. Conf. Countdown Wireless 21st Century*, vol. 2. Jul. 1995, pp. 815–819.
[41] J. Segura, "Sharp bounds for cumulative distribution functions," *J. Math. Anal. Appl.*, vol. 436, no. 2, pp. 748–763, 2016.

SARATH GOPI received the B.Tech. degree in electronics and communication engineering from the National Institute of Technology, Kozhikode, India, in 2002, and the M.Tech. in electrical engineering from IIT Bombay, India in 2008. He is currently pursuing the Ph.D. degree as an external registrant from DRDO at the Department of Electrical Engineering, IIT Madras. Since 2003, he has been a Scientist with Naval Physical and Oceanographic Laboratory, DRDO, Kochi, India. His research interests include OFDM, index modulation, and statistical signal processing.

SHEETAL KALYANI received the B.E. degree in electronics and communication engineering from Sardar Patel University, Gujarat, India, in 2002, and the Ph.D. degree in electrical engineering from the Indian Institute of Technology, Madras, India, in 2008. She was a Senior Research Engineer with the Centre of Excellence in Wireless Technology, Chennai, India, from 2008 to 2012, where she is currently an Associate Professor with the Department of Electrical Engineering. Her current research interests include generalized fading models, hypergeometric functions, performance analysis of wireless systems/networks, compressed sensing, and machine learning and reinforcement learning for wireless applications.

LAJOS HANZO received the 5-year degree in electronics and the Ph.D. degree from the Technical University of Budapest, in 1976 and 1983, respectively. In 2016, he was admitted to the Hungarian Academy of Science. During his 40-year career in telecommunications, he has held various research and academic posts in Hungary, Germany, and U.K. Since 1986, he has been with the School of Electronics and Computer Science, University of Southampton, U.K. He is currently the Chair in telecommunications with the University of Southampton. He is also a former Chaired Professor with Tsinghua University, Beijing. He has successfully supervised 119 Ph.D. students, coauthored 18 John Wiley/IEEE Press books on mobile radio communications totaling in excess of 10 000 pages, published more than 1800 research contributions at IEEE Xplore. He is currently directing 60-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC), U.K. He is a Fellow of Royal Academy of Engineering, IET, and EURASIP. In 2009, he received an Honorary Doctorate by the Technical University of Budapest and in 2015 by The University of Edinburgh. He received a number of distinctions, the European Research Councilâ Źs Advanced Fellow Grant, and the Royal Societyâ Źs Wolfson Research Merit Award. He acted both as the TPC and the General Chair of IEEE conferences and presented keynote lectures. He is a former Editor-in-Chief of the IEEE Press. He is an Enthusiastic Supporter of industrial and academic liaison. He offers a range of industrial courses. He is also a Governor of the IEEE ComSoc and VTS.