Enhanced BICM Receiver Metrics for Ultra-Reliable Low-Latency Short Block Channel Communications

Mody Sy, *Member IEEE* and Raymond Knopp, *Fellow IEEE*

***Abstract*—This paper presents bit-interleaved coded modulation metrics for joint estimation and detection in short block length channels, addressing scenarios with unknown channel state information and low training dimension density. We show that it is possible to enhance the performance and sensitivity through joint detection-estimation compared to standard receivers, especially when the channel state information is unknown and the density of the training dimensions is low. The performance analysis makes use of a full 5G transmitter and receiver chains for both Polar and LDPC coded transmissions paired with QPSK modulation scheme. We consider transmissions where reference signals are interleaved with data and both are transmitted over a small number of OFDM symbols so that near-perfect channel estimation cannot be achieved. This is particularly adapted to mini-slot transmissions for ultra-reliable, low-latency communications (URLLC) or for short packet random access use cases. Performance evaluation spans SIMO and SU-MIMO configurations, emphasizing the efficacy of BICM detection in realistic base station receiver scenarios. Our findings demonstrate that when the detection windows used in the metric units are on the order of four modulated symbols, the proposed receiver metrics can be used to achieve detection performance that is close to that of a coherent receiver with perfect channel state information.**

***Index Terms*—Bit-Interleaved Coded Modulation, 5G NR Polar code, 5G NR LDPC Code, 5G NR Physical Uplink Channels, Short Data Transmission, Unknown Channel State Information, Joint Estimation and Detection.**

I. Introduction

IT IS EXPECTED that the 6G air-interface will build upon the 5G standard and address new paradigms for feedback-based cyber-physical systems combining communications and sensing. In particular, there will be a need for tight control loops using the air-interface to control 6G-enabled objects with high-reliability, perhaps even requiring lower latencies than those achieved by current 5G technology, for example sub-1ms uplink application-layer latency in microwave spectrum. Although 5G transmission formats can provide very short-packet transmission through the use of mini-slots, the ratio of training information to data is not necessarily adapted to extremely short data transmission. Moreover, the transmission formats are designed with conventional quasi-coherent receivers, which can be quite suboptimal in such scenarios where accurate channel estimation is impossible because of sporadic transmission of short packets. One such instance is because of stringent decoding latency constraints, such as those emerging in so-called *ultra-reliable-low-latency communication* (URLLC) industrial IoT applications. This would be similar for evolved channel state information (CSI) feedback control channels or future combined-sensing and communication paradigms requiring rapid sensory feedback to the network. One of the main applications is in the field of mission-critical communications, such as those used by emergency services or in industrial control systems that require extremely high levels of reliability and low latencies.

In this work, we investigate *bit-interleaved coded modulation* (BICM) and detection strategies for packets in the range of 20–100 bits for these envisaged beyond 5G/6G signaling scenarios. Furthermore, BICM stands as a ubiquitous coding paradigm in wireless communication channels, serving as a cornerstone for contemporary high spectral efficiency systems and low spectral efficiency orthogonal modulation systems. The significance of BICM becomes especially evident in scenarios marked by error-prone communication channels, necessitating a heightened level of reliability. Its efficacy depends on detection and decoding metrics, requiring a nuanced equilibrium between enhanced performance and low complexity, especially in URLLC scenarios. Noteworthy is the historical integration of BICM into 3GPP systems, a practice dating back to the 3G-era.

A. Relevant prior of art

There is a wealth of literature on BICM receivers from various perspectives [1]–[3] demonstrating their potential impact and importance.Among the pioneers who sparked interest in BICM was the seminal work conducted by Caire *et al.* [3], wherein they provided a comprehensive analysis of BICM in terms of information rate and probability of error, encompassing both coherent and non-coherent detection. We are particularly interested in BICM *multiple input multiple output* (MIMO) receivers for joint estimation and detection. This is particularly relevant as BICM MIMO OFDM emerges as an appealing prospect for future wireless networks, wherein MIMO enhances spectrum efficiency, OFDM reduces equalization complexity, and BICM provides reliable coded-modulations. In the early 21st century, noteworthy advancements were put forth in the design of maximum likelihood receivers tailored for MIMO systems [4]–[6]. Afterwards, numerous research inquiries have been directed towards the design of low-complexity receivers for BICM MIMO, with a particular focus on low-dimensional and high-dimensional MIMO systems, but primarily restricted to coherent communication and more recently, particular attention has been paid to machine-learning-based MIMO receiver designs [7], [8]. Upon revisiting the core of this investigation, namely

Manuscript prepared on April 01, 2024. The authors are with the Department of Communication Systems, EURECOM, BIOT, 06410, France (e-mails: mody.sy@eurecom.fr).

This article was in part presented at the 2024 IEEE Conference on Standards for Communications and Networking (Munich, Germany, November 2023) and the 2024 European Conference on Networks and Communications & 6G Summit (Antwerp, Belgium, June 2024).

the transmission of short packets, it becomes apparent that this area has attracted noteworthy scholarly interest in recent years. Considerable research efforts have been dedicated to various facets, including the design of signal codes, enhanced receiver algorithms [9]–[15], as well as establishing state-of-the-art converse and achievability bounds for coherent and non-coherent communications [16]–[21].

B. Contributions

This work stands out from prior literature by introducing a novel BICM receiver design within the imperfect *channel state information* (CSI) scenario, aiming to assess the impact of various channel conditions. Hence, we present enhanced receiver metrics for short data in the range of $20-100$ bits for the envisaged beyond 5G/6G signaling scenarios by evaluating their performance over 5G short block channels, utilizing Polar and *low-density parity-check* (LDPC) coded formats. We look into receiver metrics exploiting *joint estimation and detection* (JED) which is amenable to situations where low-density *demodulation reference signals* (DMRS) are interleaved with coded data symbols. We specifically address situations where accurate channel estimation is impossible, demonstrating that a well-conceived metric exploiting interleaved DMRS in the detection metric computation achieves performance comparable to a receiver with perfect channel state information. Remarkably, this approach demonstrates substantial performance gains when compared to conventional 5G *orthogonal frequency division multiplexing* (OFDM) receivers, applicable to both uplink and downlink transmission scenarios. The proposed scheme performs detection over contiguous groups of modulated symbols, including those from the DMRS to provide soft metrics for the bits in each set, to the channel decoder. Explicitly, our main proposal consists in the non-coherent metric design/use in which channel estimation based on averaging/smoothing over the number of dimensions exhibiting channel coherence, constitutes a part of the metric for generating the *log likelihood ratio* (LLR) outputs.

Our contributions span the following principal avenues. Initially, we introduce BICM metrics tailored for non-coherent fading channels in a *single-input multiple-output* (SIMO) transmission. These metrics effectively address challenges arising from both line-of-sight (LOS) and *non-line-of-sight* (NLOS) fading channels. Subsequently, we broadened the scope of BICM metrics application within a $(\mathsf{N_R} \times \mathsf{N_T})$ *single-user* (SU)-MIMO system, specifically considering scenarios involving block fading channels. In the third phase, we proposed BICM metrics designed for a $(\mathsf{N_R} \times 2)$ SU-MIMO configuration, specifically tailored for LOS channels. Collectively, these contributions propel the state-of-the-art, pushing the boundaries in receiver algorithm designs for beyond 5G/6G short data communication systems.

C. Paper Outline

The article is structured as follows. Section II lays out the system model and the foundations of 5G polar and LDPC coded modulations, Section III focuses on the proposed BICM receiver metrics, Section IV presents the numerical results and performance analysis, and finally Section V concludes the paper.

Notation : Scalars are denoted by italic letters, vectors and matrices are denoted by bold-face lower-case and upper-case letters, respectively. For a complex-valued vector $\mathbf{x}$, $||\mathbf{x}||$ denotes its Euclidean norm, $|\cdot|$ denotes the absolute value. $\|\cdot\|_{\mathsf{F}}$ is the Frobenius norm of matrix. $\mathrm{tr}\{\cdot\}$ denotes the trace of matrix. $\mathbb{E}\{\cdot\}$ denotes the statistical expectation. $\mathrm{Re}(\cdot)$ denotes the real part of a complex number. $\mathrm{I}_0(\cdot)$ is the zero-th order modified Bessel function of the first kind. $\mathbf{I}$ is an identity matrix with appropriate dimensions. Galois field is denoted by $\mathbb{F}_2$. $\mathbf{x} \in \chi_b^j = \{\mathbf{x} : e_j = b\}$ is the subset of symbols $\{\mathbf{x}\}$ for which the $j-th$ bit of the label e is equal to $b = \{0, 1\}$. The number of bits required to a symbol is denoted by $m \triangleq \log_2(\mathcal{M})$. The cardinality of χ is given by $\mathcal{M} \triangleq |\chi|$. $\Lambda^j(\cdot)$ denotes log likelihood ratio, with $j = 1, 2, \ldots, m$. The superscripts $^{\mathsf{T}}$ and † denote the transpose and the complex conjugate transpose or Hermitian.

II. General Framework

A. Bit-Interleaved Polar-coded Modulation (BIPCM)

Bit interleaved polar coded modulation is referred to as BIPCM. In this instance, we are dealing with the *cyclic redundancy check* (CRC)-aided polar coding scheme, one of the basic code construction techniques established by the 3GPP Standard [22]. Using polar codes as a channel coding scheme for 5G control channels has demonstrated the significance of Arikan's invention [23], and its applicability in commercial systems has been proven. This new coding family achieves capacity rather than merely approaching it as it is based on the idea of channel polarization. Polar codes can be used for any code rate and for any code length shorter than the maximum code length due to their adaptability. They are the first kind of forward error correction codes that achieve symmetric capacity for any binary-input discrete memoryless channel under low-complexity encoding and low-complexity *successive cancellation* (SC) decoding with quasi-linear order for infinite length codes. In 5G new radio, the polar codes are employed to encode *broadcast channel* (BCH) as well as *downlink control information* (DCI) and *uplink control information* (UCI). Furthermore, the transmission process is straightforward and complies with the 3GPP standard specifications [22].

Assume that the input message (*UL/DL Control Information*) before CRC attachment is $a(0), a(1), \ldots, a(A-1)$, where A is input sequence, parity bits are $p(0), p(1), \ldots, p(L-1)$, L is the number of parity bits. For the downlink, a CRC of length $L = 24$ bits is used, and for the uplink, depending on the quantity of A, CRCs of length $L = 6$ and $L = 11$ bits are used. The message bits after attaching CRC are $b(0), b(1), \ldots, b(B-1)$, where B is the size of control information with CRC bits: $B = A + L$. The input bit sequence to the code block segmentation is denoted $a(0), a(1), \ldots, a(A-1)$, where the value of A is no larger than 1706. Assume that the maximum code block size is A' and C the number of existing code blocks, the sequence $c_r(0), c_r(1), \ldots, c_r(A'/C - 1)$ is used to calculate the CRC parity bits $p_r(0), p_r(1), \ldots, p_r(L-1)$. The sequence of bits resulting after attaching a CRC to the $r-th$ code block is denoted by $c_r(0), c_r(1), \ldots, c_r(K_r - 1)$,

where K_r is the number of bits in the $r-th$ code block to be fed to the channel encoder. Then, the coded bit are denoted by $d_r(0), d_r(1), \ldots, d_r(N_r-1)$ where $N_r = 2^n$. Denote by E_r the rate matching output sequence length of the $r-th$ coded-block: *if* $E_r \leq (9/8) \cdot 2^{(\lceil \log_2 E_r \rceil - 1)}$ and $K_r/E_r < 9/16$, $n_1 = \lceil \log_2 E_r \rceil - 1$, *else* $n_1 = \lceil \log_2 E_r \rceil$.
And then, $R_{\min} = 1/8$; $n_2 = \lceil \log_2 (K/R_{\text{min}}) \rceil$; $n = \max\{\min\{n_1, n_2, n_{\max}\}, n_{\text{min}}\}$) where $n_{\min}$ and $n_{\max}$ provide a lower and an upper bound on the code length, respectively. In particular, and $n_{\min} = 5$ and $n_{\max} = 9$ for the downlink control channel, whereas $n_{\max} = 10$ for the uplink control channel. The polar encoding process is based on a well-defined procedure outlined in [22]:

1) The sequence $c_r(0), c_r(1), \ldots, c_r(K_r-1)$ is interleaved into bit sequence $c'_r(0), c'_r(1), \ldots, c'(K_r-1)$ via a definite interleaving pattern [22].
2) The interleaved vector $\mathbf{c}'$ is assigned to the information set along with the *polar codes* PC bits, while the remaining bits in the N_r-bit vector $\mathbf{u}$ are frozen. Hence, $\mathbf{u}_r = u_r(0), u_r(1), \ldots, u_r(N_r-1)$ is generated according to the clause 5.3.1.2 [22].
3) Denote $\mathbf{G}_{N_r} = (\mathbf{G}_2)^{\otimes n}$ as the $n-th$ Kronecker power of matrix $\mathbf{G}_2$, where $\mathbf{G}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, the output after encoding $\mathbf{d}_r = d_r(0), d_r(1), \ldots, d_r(N_r-1)$, such that $\mathbf{d}_r = \mathbf{u}_r \mathbf{G}_{N_r} \in \mathbb{F}_2$.

Subsequently, a rate matching process is performed per coded block, involving sub-block interleaving, bit collection, and bit interleaving. The output bit sequence after rate matching consists of sequences $e_{r,k}$, where $r = 0, \ldots, C-1$ and $k = 0, \ldots, E_r - 1$, with E_r being the number of rate-matched bits for the r-th code block. Different techniques such as perforation, shortening, or repetition ($E_r \geq N_r$) may be applied during rate matching to convert a vector of N_r bits into a vector of E_r bits. Following rate matching, code block concatenation is performed to convert all code block messages into a sequence of transport block messages. The output bit sequence after code block concatenation is denoted as $g(0), g(1), \ldots, g(E-1)$. Regarding the decoding process, several main polar code decoding algorithms are currently used, including the SC algorithm [23], the *successive cancellation list* (SCL) algorithm [24], [25], the *CRC-aided SCL* (CA-SCL) algorithm [26], [27], the *belief propagation* (BP) algorithm [28], and the *successive cancellation with adaptive node* (SCAN) algorithm [29]. The SCL algorithm improves upon the SC algorithm by providing multiple paths and outperforms it in terms of performance. The CA-SCL algorithm incorporates a high-rate CRC code to assist in selecting the correct codeword from the final list of paths in the SCL decoder, effectively enhancing its reliability. It has been observed that the right codeword is usually included in the list every time the SCL decoder fails.
The performance ranking of the decoding algorithms is as follows: `CA-SCL> state-of-the-art SCL>BP=SCAN>SC`. Therefore, for improved performance, the channel decoder technique should utilize CA-SCL decoding for downlink (DCI or BCH) or uplink (UCI) messages. The adoption of polar codes by 3GPP was partly due to the well-acknowledged potential of CA-SCL decoding to outperform Turbo or LDPC codes.
Figure 1 presents an overview of the BIPCM/BILCM process short block uplink channels. The transmit-end procedure includes several steps such as adding a transport block CRC, segmenting code blocks with additional CRC attachment, channel encoding, rate matching, code block concatenation, and modulation. It is important to emphasize that the receiving chain simply follows the reverse flow of the transmitting chain.

B. Bit-Interleaved LDPC-coded Modulation (BILCM)

Bit-Interleaved LDPC-Coded Modulation is referred to as BILCM. First proposed by Gallager in the early 1960s [30], LDPC coding has proven to be highly suitable for 5G NR due to its advantages such as high throughput, low latency, efficient decoding complexity, and rate compatibility. The performance of LDPC codes in 5G NR is impressive, exhibiting an error floor at or below the 10^{-5} *block error rate* (BLER), a significant improvement over traditional coding techniques. The construction of an NR LDPC code involves a base graph matrix ($\mathbf{BG}$) of dimension $M \times N$, denoted as $\mathbf{H}_{\mathsf{BG}}$. The choice of $\mathbf{H}_{\mathsf{BG}}$ matrices in the 5G NR coding process depends on the coding rate and the length of the transport block or code block. Two base graphs are commonly used: $\mathbf{BG}_1$ with dimensions $N = 68$ and $M = 46$, optimized for large information block sizes ($K \leq 8448$) and high coding rates ($1/3 \leq R \leq 8/9$), and $\mathbf{BG}_2$ with dimensions $N = 52$ and $M = 42$, optimized for small information block sizes ($K \leq 3840$) and lower coding rates ($1/5 \leq R \leq 2/3$). These LDPC codes are particularly well-suited for scenarios requiring high reliability, as they offer additional coding gain at low code rates. The maximum number of information bits for $\mathbf{BG}_1$ is $K = 22Z_c$, and for $\mathbf{BG}_2$ it is $K = 10Z_c$, where Z_c represents the lifting size. Each base graph has 51 different lifting sizes, ranging from 2 to 384. The parity check matrix $\mathbf{H}$ is derived from $\mathbf{H}_{\mathsf{BG}}$ by replacing each element with a cyclic permutation identity matrix, denoted as $\mathbf{I}(\mathrm{p_{ij}})$. In this process, each element of $\mathbf{H}_{\mathsf{BG}}$ is substituted with the corresponding *cyclic permutation matrix* (CPM), along $\mathrm{p_{ij}}$ times. The resulting matrix $\mathbf{H}$ has a size of $m \times n$, with $m = M \times Z_c$, $n = N \times Z_c$, and $k = n - m = (N-M) \times Z_c$. Both $\mathbf{BG}_1$ and $\mathbf{BG}_2$ share similar structures.

Due to the specific structure and features of these base graphs, various effective LDPC encoding techniques have been proposed in the literature. Recently, a novel and efficient encoding technique was proposed in [31], which offers high throughput and low complexity, making it an attractive option for LDPC coding architecture. Furthermore, the BILCM transmission procedure is almost identical to that described with BIPCM, starting by attaching a CRC to the transport block.

Let the transport message before CRC attachment be denoted as $a(0), a(1), \ldots, a(A-1)$, where A is the size of the transport block message. The parity bits are represented as $p(0), p(1), \ldots, p(L-1)$, where L is the number of parity bits, representing the CRC length. If $A > 3824$, L is set to 24, otherwise, it is set to 16. The message bits after attaching CRC are denoted as $b(0), b(1), \ldots, b(B-1)$, where B is the

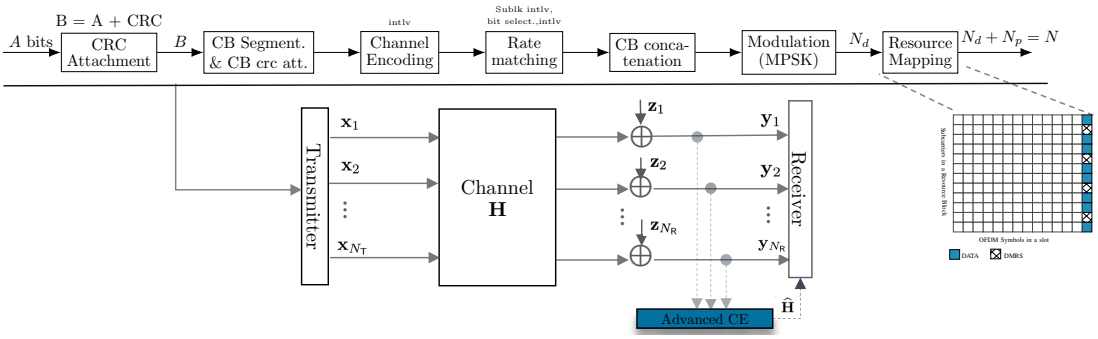


Fig. 1. Bit-Interleaved Polar/LDPC coded Modulation (BIPCM/BILCM) : Transmitter end.

size of the transport block information with CRC bits, and $B = A + L$. The LDPC base graph is selected based on the size of the transport block message A and the transport block coding rate R. If $A \leq 292$, or if $A \leq 3824$ and $R \leq 0.67$, or if $R \leq 0.25$, LDPC $\mathbf{BG}_2$ is used. Otherwise, LDPC $\mathbf{BG}_1$ is used. The output of code block segmentation and CRC attachment is denoted as $c_r(0), c_r(1), \ldots, c_r(K_r - 1)$, where $K_r = K'_r + L$, and K'_r represents the number of bits in the r-th code block, and L is the number of attached CRC bits on the r-th code block. Each code block message is encoded independently. In 3GPP NR, the input bit sequence is represented as $\mathbf{c}_r = [c(0), c(1), \ldots, c(K_r - 1)]^{\mathsf{T}}$, where K_r is the number of information bits in a code block, and the redundant bits are called parity bits, denoted by $\mathbf{w} = [w(0), w(1), \ldots, w(N_r + 2Z_c - K_r - 1)]^{\mathsf{T}}$. The LDPC coded bits are denoted by $d_r(0), d_r(1), \ldots, d_r(N_r - 1)$.

A code block is encoded by the LDPC encoder based on the following procedure [22]:

1) Find the set with index iLS which contains Z_c in [22].
2) Set $d_{r,k-2Z_c} = c_k, \forall k = 2Z_c, \ldots, K_r - 1$
3) Generate $N_r + 2Z_c - K_r$ parity bits $\mathbf{w} = [w(0), w(1), \ldots, w(N_r + 2Z_c - K_r - 1)]^T$ such that $\mathbf{H} \times [\mathbf{c}_r \ \mathbf{w}]^{\mathsf{T}} = \mathbf{0}$
4) The encoding is performed in $\mathbb{F}_2$.
5) Set $d_{r,k-2Z_c} = w_{k-K_r}, \forall k = K_r, \ldots, N_r + 2Z_c - 1$

The subsequent steps involve rate matching and code block concatenation. At the receiver, the LDPC decoding is performed on each code block individually. For LDPC decoding, various techniques can be implemented, with belief propagation (BP) methods being the most commonly used. BP methods rely on iterative message exchange between bit nodes and check nodes, offering near-optimal decoding performance at the cost of computational complexity. However, to strike a better balance between performance and complexity, several simplified and effective decoding algorithms have been proposed in the scientific literature. One such decoding algorithm is layered message passing [32], which stands out as a promising approach for URLLC due to its ability to speed up convergence times [33], [34], making it a suitable candidate for short packet transmissions.

C. Modulation and Resource Mapping

In both scenarios, the encoded payload undergoes rate-matching and code block concatenation prior to being fed into a QPSK modulator. This process yields a set of complex-valued modulation symbols. Subsequently, the resource allocation process is executed, where one or multiple OFDM symbols are used to allocate the modulated symbols to resource blocks and insert the DMRS resources. The number of resource blocks is determined by the payload size and coding settings. When the payload size is small, fewer resource blocks are required, thereby maintaining a constant effective coding rate. As illustrated in Figure 2, the resource mapping process in use is in the same spirit as the 3GPP *physical uplink control channel* (PUCCH) format 2 transmission [35].

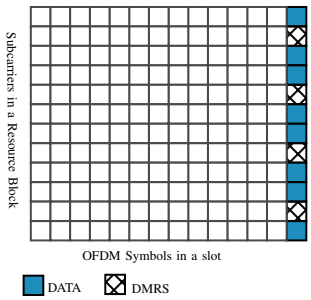


Fig. 2. General resource mapping: 1 OFDM symbol.

Furthermore, the transmitted signal $\mathbf{x}$ typically consists of data-dependent $\mathbf{x}^{(\mathsf{d})}$ and data-independent $\mathbf{x}^{(\mathsf{p})}$ components, known as pilot or reference signals. The reference signals are used to resolve channel ambiguity in time, frequency and space. Specifically, they are employed to estimate the channel. In practice, the reference signals are commonly interleaved among the data-dependent components. It is notably the case in current OFDM systems. In earlier CDMA systems, reference signals were sometimes superimposed on top of data-dependent signals. The number of data dimensions is denoted by N_d, and the number of reference signal dimensions is denoted by N_p, where $N_d + N_p = N$. In 3GPP standard, N is typically equal to $12\mathcal{K}\mathcal{L}$. This represents the number of complex dimensions or resource elements(REs) in the physical resource blocks (PRBs). The number of PRBs, $\mathcal{K}$ ranges from 1 to 16, while the number of OFDM symbols, $\mathcal{L}$, ranges from 1 to 14, and can be increased if multiple slots are used for signaling the channel bits. The assumption in this work is that the data-dependent components of $\mathbf{x}$ are generated from a binary code whose output is interleaved and mapped to an $\mathcal{M}$-ary modulation symbol alphabet. We will assume that the binary code generates E bits and the interleaver mapping is one-to-one, so that E bits are also fed to the modulator.

The binary-code and interleaver combination can thus be seen as a (E, B) binary block code. Denote the E coded bits as $e_k, k = 0, 1, \cdots, E-1$. Adjacent $\log_2 \mathcal{M}$ bit-tuples are used to select the modulated symbols in the symbol alphabet. Typically, we will assume that a Gray mapping is used in the case of non-binary modulation.

D. Pilot allocation procedure within the spatial dimension

The resources mapping procedure within spatially multiplexed MIMO systems. The resource mapping procedure in spatially multiplexed MIMO systems needs to consider the way in which pilot symbols are allocated in the spatial dimension, in addition to the time and/or frequency dimensions. Precisely, the transmission of training symbols must be executed in a manner that prevents interference, ensuring the accurate estimation of channel state information. This consideration involves the exploration of pilot allocation aspects that leverage frequency orthogonality, time orthogonality, and signal orthogonality.The most straightforward method for interleaving pilot symbols and data is arguably through time and/or frequency insertion, although other approaches such as superposition or code-division multiplexing are also possible [36]. In the instance of block fading, at least $\mathsf{N_T}$ pilot symbols must be inserted in each coherence block, one per antenna, with $N_\mathsf{p} \geq \mathsf{N_T}$ pilot symbols per coherence block [37]. Widely adopted in today's emerging communication standards, frequency orthogonality seems to have more merit and is particularly well suited to OFDM-type systems.

III. BICM Receiver Design

A. A $(\mathsf{N_R} \times 1)$ SIMO within Non-Coherent Fading Channel

Considering a SIMO OFDM BICM system with a single antenna element on the transmit array ($\mathsf{N_T} = 1$) and multiple element receive arrays ($\mathsf{N_R}$). The transmitted and received signals are N-dimensional column vectors, and thus a system is designed in such a way that the relationship between the transmitted and received signals is as follows: $\mathbf{y}_i = \mathbf{h}_i \mathbf{x} + \mathbf{z}_i, \quad i = 0, 1, \cdots, \mathsf{N_R} - 1$, where $\mathbf{z}_i$ is additive white Gaussian noise whose real and imaginary components are independent and have variance and $\mathbf{h}_i$ represents the channel vector.

1) Perfect Channel State Information

In the instance of perfect channel state information, the likelihood function is shown to be:

$$\mathrm{q}\left(\mathbf{x}, \{\mathbf{y}_i, \mathbf{h}_i\}\right) = \mathrm{p}\left(\{\mathbf{y}_i, \mathbf{h}_i\} \mid \mathbf{x}\right) = \mathrm{p}\left(\{\mathbf{y}_i\} \mid \mathbf{x}, \mathbf{h}_i\right) \mathrm{p}\left(\{\mathbf{h}_i\} \mid \mathbf{x}\right) \tag{1}$$

If the transmitted signal $\mathbf{x}$ is independent of the channel realization $\{\mathbf{h}_i\}$, the term $p\left(\{\mathbf{h}_i\} \mid \mathbf{x}\right)$ in (1) can be dropped since it will disappear in (2). The likelihood function is commonly equivalent to :

$$\mathrm{q}\left(\mathbf{x}, \{\mathbf{y}_i, \mathbf{h}_i\}\right) \propto \exp\left(-\frac{1}{\mathrm{N}_0} ||\mathbf{y}_i - \mathbf{h}_i \mathbf{x}||^2\right). \tag{2}$$

Using the norm extension property, ignoring terms that are independent of $\mathbf{x}$, the likelihood function then simply becomes:

$$\mathrm{q}\left(\mathbf{x}, \{\mathbf{y}_i, \mathbf{h}_i\}\right) \propto \quad \exp\left(\frac{2}{\mathrm{N}_0} \mathrm{Re}\left(\mathbf{y}_i \mathbf{h}_i^\dagger \mathbf{x}^\dagger\right) - \frac{1}{\mathrm{N}_0} ||\mathbf{h}_i \mathbf{x}||^2\right). \tag{3}$$

The likelihood of coded bit $e_j \in \{0, 1\}$ is

$$\mathrm{q}\left(e_j(\mathbf{x}) = b, \{\mathbf{y}_i, \mathbf{h}_i\}\right) = \sum_{\mathbf{x} \in \chi_b^j} \mathrm{q}\left(\mathbf{x}, \{\mathbf{y}_i, \mathbf{h}_i\}\right). \tag{4}$$

As is common in the case of BICM-based systems, the soft input to the binary channel decoder is given as the log-likelihood ratio (LLR) for the $j-th$ coded bit, such that :

$$\Lambda^j\left(\mathbf{y}_i\right) = \log \frac{\mathrm{q}\left(e_j(\mathbf{x}) = 0, \{\mathbf{y}_i, \mathbf{h}_i\}\right)}{\mathrm{q}\left(e_j(\mathbf{x}) = 1, \{\mathbf{y}_i, \mathbf{h}_i\}\right)}. \tag{5}$$

We simplify (5) using a *max-log approximation*: $\log\left\{\sum_i \exp\left(\lambda_i\right)\right\} \sim \max_i \left\{\lambda_i\right\}$, resulting in (6)

$$\begin{aligned} \Lambda^j\left(\mathbf{y}\right) = & \max_{\mathbf{x} \in \chi_0^j} \frac{1}{\mathrm{N}_0} \left(\sum_{i=0}^{\mathsf{N_R}-1} 2\,\mathrm{Re}\left(\mathbf{y}_i \mathbf{h}_i^\dagger \mathbf{x}^\dagger\right) - ||\mathbf{h}_i \mathbf{x}||^2\right) \\ & - \max_{\mathbf{x} \in \chi_1^j} \frac{1}{\mathrm{N}_0} \left(\sum_{i=0}^{\mathsf{N_R}-1} 2\,\mathrm{Re}\left(\mathbf{y}_i \mathbf{h}_i^\dagger \mathbf{x}^\dagger\right) - ||\mathbf{h}_i \mathbf{x}||^2\right). \end{aligned} \tag{6}$$

We consider the ideal receiver, referred to as *Perfect CSI*, as a benchmark for comparison with the subsequent receiver architectures. This above metric is typically employ in conventional quasi-coherent receivers based on a separate least-squares channel estimation method by substituting $\mathbf{h}$, with the corresponding estimated channel $\hat{\mathbf{h}}$. Moreover, within the framework of a conventional receiver, it is presupposed that, at the very least, the observation of a single reference signal spans the entirety of a *physical resource block* (PRB) to generate the coded bits corresponding to each data symbol within that PRB. Consequently, a block is construed as comprising a singular data symbol assisted at least by one reference symbol dimension. We will note this case throughout this manuscript as *No CSI* ($N_d = 1$).

2) Unknown Channel State information

We describe BICM metrics for a general non-coherent fading channel with unknown phase on the line-of-sight (LOS) components and fully unknown diffuse (Non-LOS) components. The overall unknown channel gain is given by $\mathbf{h}_i = \left(\sqrt{\alpha} e^{j\theta_i} + \sqrt{1-\alpha}\mathrm{h}_i^{(f)}\right)\mathbf{I}$, where θ_i is assumed to be i.i.d. uniform random variables on $[0, 2\pi)$, $\mathrm{h}_i^{(f)}$ is a zero-mean, unit-variance, circularly-symmetric complex Gaussian random variable and α is the relative strength of the LOS component. The amplitude $|\mathrm{h}_i|$ on each receiver is thus Ricean distributed. It is worth noting that the i.i.d. assumption for the θ_i is somewhat unrealistic for a modern array receiver with accurate calibration. The phase differences would be more appropriately characterized by two random-phases, one originating from the time-delay between transmitter and receiver and the other from the angle of arrival of the incoming wave. The phase differences of individual antenna elements for a given carrier frequency could then be determined from the angle of arrival and the particular geometry of the array. To avoid assuming a particular array geometry, the i.i.d. uniform model provides a simpler and universal means to derive a receiver metric.

$$\Lambda^{j}(\mathbf{y}) = \max_{\mathbf{x} \in \chi_0^j} \left(\sum_{i=0}^{\mathsf{N_R}-1} -\frac{\alpha \|\mathbf{x}\|^2}{\mathbf{L}_{\mathsf{x}}} + \beta_x \left|\mathbf{x}^{\dagger}\mathbf{y}_i\right|^2 + \frac{2\sqrt{\alpha}}{\mathbf{L}_{\mathsf{x}}} \left|\mathbf{x}^{\dagger}\mathbf{y}_i\right| \right) - \sum_{\mathbf{x} \in \chi_0^j} \mathsf{N_R} \log\left(\mathbf{L}_{\mathsf{x}}\right) - \max_{\mathbf{x} \in \chi_1^j} \left(\sum_{i=0}^{\mathsf{N_R}-1} -\frac{\alpha \|\mathbf{x}\|^2}{\mathbf{L}_{\mathsf{x}}} + \beta_x \left|\mathbf{x}^{\dagger}\mathbf{y}_i\right|^2 + \frac{2\sqrt{\alpha}}{\mathbf{L}_{\mathsf{x}}} \left|\mathbf{x}^{\dagger}\mathbf{y}_i\right| \right) + \sum_{\mathbf{x} \in \chi_1^j} \mathsf{N_R} \log\left(\mathbf{L}_{\mathsf{x}}\right). \tag{8}$$

Proposition 1. After neglecting multiplicative terms independent of the transmitted message, the likelihood function can be expressed as follows:

$$\mathrm{q}(\mathbf{x},\mathbf{y}) = \prod_{i=0}^{\mathsf{N_R}-1} \frac{1}{\mathbf{L}_{\mathsf{x}}} \exp\left(-\frac{\alpha \|\mathbf{x}\|^2}{\mathbf{L}_{\mathsf{x}}} + \beta_x \left|\mathbf{x}^{\dagger}\mathbf{y}_i\right|^2 \right) \times \mathrm{I}_0\left(\frac{2\sqrt{\alpha}}{\mathbf{L}_{\mathsf{x}}} \left|\mathbf{x}^{\dagger}\mathbf{y}_i\right| \right), \tag{7}$$

where $\mathbf{L}_{\mathsf{x}} = \mathrm{N}_0 + 2(1-\alpha)\|\mathbf{x}\|^2$, $\beta_x = \frac{2(1-\alpha)}{\mathrm{N}_0(\mathrm{N}_0+2(1-\alpha)\|\mathbf{x}\|^2)}$ and $\mathrm{I}_0(\cdot)$ is the zero-order modified Bessel function

Proof. see appendix section A.

Then we succinctly use (4-5) to generate the LLR of the j-th coded bit.

Note that in the above expressions, we do not limit the dimensionality of the observations when computing likelihoods of particular bits. In the original work of Caire *et al.* [3], the authors assume an ideal interleaving model which allows limiting the observation interval of a particular coded bit to the symbol in which it is conveyed. For long blocks, this assumption is realistic for arbitrary modulation signal sets and is sufficient for BPSK and QPSK irrespective of the block length when the channel is known perfectly. Nevertheless, practical systems usually apply single symbol likelihood functions for short blocks and high-order modulations. Furthermore, for the primary case of interest here, namely transmission without channel state information, single symbol detection is impossible. At the very least, the observation of one reference symbol must be used to generate likelihoods of the coded bits of a data symbol, thus warranting the study of block detection.

Corollary 1. The LLR binary metric calculations based on (7) in the logarithmic domain can be computationally complex to implement. To simplify the calculations, the *max-log approximation* is commonly used. First, an exponential approximation is applied to the modified Bessel function of the first kind $\mathrm{I}_0(z)$, which results in the approximation $\mathrm{I}_0(z) \sim \frac{e^z}{\sqrt{2\pi z}} \sim e^z$.

Using this approximation, the log-likelihood ratio (LLR) for coded bit j is given in (8).

Remark 1. Furthermore, in equation (8), many terms can be omitted when the magnitude of vector $\mathbf{x}$ remains constant, as is the scenario in BPSK or QPSK modulation, for example. Additionally, in the presence of strong line-of-sight (LOS) channels, the quadratic terms in equation (8) can also be disregarded. Conversely, the computational complexity of the LLR metric in a BICM SIMO system is typically on the order of $\mathcal{O}(\mathsf{N_R} \times N_d \times \log_2 \mathcal{M})$. This complexity exhibits linearity with respect to the number of received data symbols (N_d), the number of receiving ($\mathsf{N_R}$) antennas , as well as the size of the modulation alphabet ($\mathcal{M}$). Consequently, an increase in the number of symbols or receiving antennas results in a linear escalation of computational complexity.

3) Joint Estimation and Detection

For the case of polar or LDPC-coded data, there is a compelling motivation to divide the coded streams into smaller blocks for detection due to complexity reasons. Furthermore, assuming an ideal interleaving scenario with known channels [3], detection can be performed on individual modulated symbols. However, in the presence of joint detection and estimation, where interleaved DMRS and data symbols are considered, we need to deal with short blocks that encompass both data and DMRS symbols. To achieve this, the N-dimensional vectors $\mathbf{y}$ and $\mathbf{x}$ are subdivided into smaller, more manageable segments of blocks. Subsequently, the detection metric is applied to each of these underlying segments.

Proposition 2. Observing the structure of the metrics and the absence of overlap between the data and DMRS symbols, we can easily see that the estimated *channel impulse response* (CIR) is part of the metric. By writing $\mathbf{x} = \mathbf{x}^{(\mathsf{p})} + \mathbf{x}^{(\mathsf{d})}$ where d and p are subscripts representing data, DMRS components, respectively, we can reveal $\hat{\mathbf{h}}^{\mathrm{LS}}$ in the metrics:

$$\left|\mathbf{x}^{\dagger}\mathbf{y}_i\right| = \left| \underbrace{\mathbf{x}^{(\mathsf{p})\dagger}\mathbf{y}_i^{(\mathsf{p})}}_{\text{channel estimate}} + \mathbf{x}^{(\mathsf{d})\dagger}\mathbf{y}_i^{(\mathsf{d})} \right| = \left| N_p \hat{\mathrm{h}}_i^{\mathrm{LS}} + \ \mathbf{x}^{(\mathsf{d})\dagger}\mathbf{y}_i^{(\mathsf{d})} \right|. \tag{9}$$

Mathematically, we can deduce the following relationship: $\mathbf{x}^{(\mathsf{p})\dagger}\mathbf{y}_i^{(\mathsf{p})} = \left(\mathbf{x}^{(\mathsf{p})\dagger}\mathbf{x}^{(\mathsf{p})}\right)\hat{\mathrm{h}}_i^{\mathrm{LS}} = \|\mathbf{x}^{(\mathsf{p})}\|^2 \hat{\mathrm{h}}_i^{\mathrm{LS}} = N_p \rho \hat{\mathrm{h}}_i^{\mathrm{LS}}$ where N_p number of pilots and ρ is the reference signal power and is typically normalized to unity.

The channel impulse response $\hat{\mathrm{h}}_i^{\mathrm{LS}}$ is obtained via a joint least-squares (LS) channel estimation using averaging or smoothing over the number of dimensions exhibiting channel coherence, as illustrated in Figure 3. In the process of short-block detection, we can make use of such a channel estimate like that. In general, the channel estimation procedure will work as usual and the resulting estimates are fed into the metrics considered here.

Corollary 2. Considering an iterative JED for more reliable output LLRs, the estimation-detection process based on the proposition 2 can be extended to include iterative steps to enhance transmission quality.

In general, the choice of where to apply the iterative process depends on the specific system requirements, complexity, available resources, transmission channel conditions,

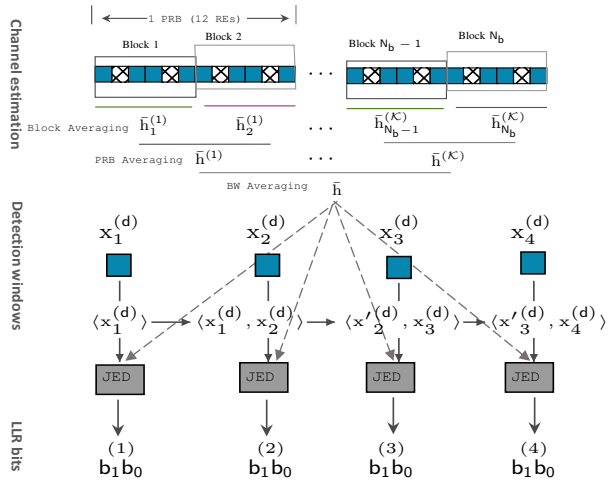


Fig. 3. Conceptual illustration of JED principle with detection windows of order $N_d = 4$.

and performance goals. Each step has different implications in terms of complexity and efficiency. Moreover, the number of iterations possible in a communication system typically depends on specific design requirements and system constraints. In the case of sporadic transmission of short packets, strict latency constraints limit the number of iterations. Therefore, the number of iterations to be implemented must be finely adjusted to strike a balance between complexity, performance, and efficiency, aiming to achieve the desired transmission quality goals. Additionally, a minimal number of iterations can be used for basic adjustments, while more complex scenarios may implement multiple iterative processes for progressive optimization.

B. A $(\mathsf{N_R} \times \mathsf{N_T})$ *MIMO Within Rayleigh Block Fading Channel*

Consider a SU-MIMO transmission model featuring multiple antenna elements in both transmitter and receiver arrays. The system dimensions are defined by parameter $\mathsf{N_R} \times \mathsf{N_T}$, $\mathsf{N_T}$ and $\mathsf{N_R}$ represent the number of transmitting and receiving antennas, respectively. We assume no *inter-symbol interference* (ISI) and consider a time-invariant configuration, making it feasible to use the standard baseband complex-valued representation. Let $\mathrm{h}_{i,j}$ represent the complex-valued path gain, serving as the fading coefficient from transmit antenna j to receive antenna i. At any given time instance, the complex-valued signals $\{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_{\mathsf{N_T}}\}$ are transmitted through the $\mathsf{N_T}$ respective antennas. The received signal $\mathbf{y}$ belongs to the complex vector space $\mathbb{C}^{1\times \mathsf{N_T}}$ while the received signal $\mathbf{y}$ belongs to the complex vector space $\mathbb{C}^{1\times \mathsf{N_R}}$. The additive white Gaussian noise $\mathbf{z}$ belongs to the complex vector space $\mathbb{C}^{1\times \mathsf{N_R}}$ with independent real and imaginary components, each having a variance of σ^2 in every dimension, and the channel matrix is denoted by $\mathbf{H} \in \mathbb{C}^{\mathsf{N_R}\times \mathsf{N_T}}$. The MIMO channel model is succinctly expressed as

$$\mathbf{y} = \mathbf{x}\mathbf{H} + \mathbf{z}. \tag{10}$$

However, the system model described in (10), which represents transmission within a single symbol interval, can be extended to accommodate the transmission of several consecutive vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ over the channel. Here, N denotes the total number of symbol intervals utilized for transmission. For the sake of clarity, we employ a matrix framework. As a result, we organize the transmitted, received, and noise vectors into matrices, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^{\mathsf{T}}$, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]^{\mathsf{T}}$, $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N]^{\mathsf{T}}$, respectively.

The elements within $\mathbf{H}$ represent the complex-valued channel gains between each transmit and receive antenna. We adopt a wide assumption regarding $\mathbf{H}$, which is that its elements, $\mathrm{h}_{\mathrm{i,j}}$, are statistically independent for the sake of simplicity. In practice, the complex path gains $\mathrm{h}_{\mathrm{i,j}}$ exhibit correlations influenced by factors such as the propagation environment, antenna element polarization, and their spacing. Accordingly, the entries of $\mathbf{H}$ can be treated as independent zero-mean complex Gaussian random variables with unit variance. This channel model is often referred to as the *identically and independently distributed* (IID) *Rayleigh fading MIMO channel model*, or more precisely, the Rayleigh block-fading.

1) Perfect Channel State Information

The channel matrix $\mathbf{H}$ is assumed to be perfectly known by the receiver. The likelihood function or conditional probability density is approximately given by:

$$\mathrm{q}\left(\mathbf{X}, \{\mathbf{Y}, \mathbf{H}\}\right) \propto \exp\left(-\tfrac{1}{\mathrm{N}_0}\,\mathrm{tr}\left\{\left(\mathbf{Y}-\mathbf{X}\mathbf{H}\right)\left(\mathbf{Y}-\mathbf{X}\mathbf{H}\right)^{\dagger}\right\}\right). \tag{11}$$

Using the norm extension property $\|\mathbf{Y}-\mathbf{X}\mathbf{H}\|_{\mathsf{F}}^2 = \|\mathbf{Y}\|_{\mathsf{F}}^2 + \|\mathbf{X}\mathbf{H}\|_{\mathsf{F}}^2 - 2\,\mathrm{Re}\left(\mathrm{tr}\left\{\mathbf{Y}\mathbf{X}^{\dagger}\mathbf{H}^{\dagger}\right\}\right)$ and neglecting terms that are independent of $\mathbf{X}$, it comes,

$$\mathrm{q}\left(\mathbf{X}, \{\mathbf{Y}, \mathbf{H}\}\right) \propto \exp\left(\tfrac{2}{\mathrm{N}_0}\,\mathrm{Re}\left(\mathrm{tr}\left\{\mathbf{Y}\mathbf{H}^{\dagger}\mathbf{X}^{\dagger}\right\}\right) - \tfrac{1}{\mathrm{N}_0}\,\|\mathbf{H}\mathbf{X}\|_{\mathsf{F}}^2\right). \tag{12}$$

The LLR bit metric for the $j-th$ bit in BICM receiver is

$$\Lambda^j\left(\mathbf{Y}\right) = \log\frac{\mathrm{q}\left(e_j(\mathbf{X})=0, \{\mathbf{Y}, \mathbf{H}\}\right)}{\mathrm{q}\left(e_j(\mathbf{X})=1, \{\mathbf{Y}, \mathbf{H}\}\right)}. \tag{13}$$

and the corresponding max-log approximation of (13) is shown to be

$$\begin{aligned}\Lambda^j\left(\mathbf{Y}\right) = &\max_{\mathbf{x}\in\chi_0^j}\frac{1}{\mathrm{N}_0}\left(2\,\mathrm{Re}\left(\mathrm{tr}\left\{\mathbf{Y}\mathbf{H}^{\dagger}\mathbf{X}^{\dagger}\right\}\right) - \|\mathbf{X}\mathbf{H}\|_{\mathsf{F}}^2\right)\\ &- \max_{\mathbf{x}\,\in\,\chi_1^j}\frac{1}{\mathrm{N}_0}\left(2\,\mathrm{Re}\left(\mathrm{tr}\left\{\mathbf{Y}\mathbf{H}^{\dagger}\mathbf{X}^{\dagger}\right\}\right) - \|\mathbf{X}\mathbf{H}\|_{\mathsf{F}}^2\right).\end{aligned} \tag{14}$$

Remark 2. We consider the ideal receiver, denoted as *Perfect CSI* to be a benchmark for comparison with the subsequent receiver architectures. These subsequent architectures typically employ a separate least-squares channel estimation method by substituting the actual channel matrix, denoted as $\mathbf{H}$, with an estimated channel matrix $\hat{\mathbf{H}}$. The separate least-squares channel estimation aims to independently estimate the channel characteristics for each antenna, disregarding inter-antenna correlation considerations.

2) Rayleigh Block Fading Channel With No CSI

As described earlier, it is assumed that the elements of the channel matrix $\mathbf{H}$ are zero- mean complex-valued Gaussian random variables with unit variance, i.e., $\mathbf{H} \sim \mathbb{C}\mathcal{N}(0, \mathbf{I})$. The complex-valued fading coefficient $\mathrm{h}_{i,j}$ represents the channel gain between the j-th transmit antenna and the i-th receive antenna. These fading coefficients are assumed to

be constant over the N symbol periods and are independent. Therefore, the magnitudes of the channel gains $\mathrm{h}_{i,j}$ have a Rayleigh distribution and their probability density function can be expressed as $\mathrm{p}\left(\mathrm{h}_{i,j}\right) = \frac{1}{\pi}\exp\left(-|\mathrm{h}_{i,j}|^2\right)$.

Both the fading coefficients and the noise follow complex Gaussian distributions. Thus, conditioned on the transmitted signal, the received signals are jointly complex Gaussian. The received signal is zero mean $\mathbb{E}\{\mathbf{Y}|\mathbf{X}\} = 0$, circularly symmetric complex Gaussian with a $N \times N$ covariance matrix $\mathbf{\Phi}_{\mathsf{Y}}$, concretely.

Definition 1. [Complex-valued Gaussian distribution]
Let $\boldsymbol{x} \in \mathbb{C}^{\mathsf{N_T}}$, then the probability density function $f_{\boldsymbol{x}}(\boldsymbol{v})$ of $\boldsymbol{x}$ is given by:

$$f_{\boldsymbol{x}}(\boldsymbol{v}) = \frac{1}{\pi \det\left(\mathbf{\Phi}_{\boldsymbol{x}}\right)} \exp\left[-\left(\boldsymbol{v}-\boldsymbol{\mu}_{\boldsymbol{x}}\right)^{\dagger} \mathbf{\Phi}_{\boldsymbol{x}}^{-1}\left(\boldsymbol{v}-\boldsymbol{\mu}_{\boldsymbol{x}}\right)\right] . \tag{15}$$

Based on the above definition, the likelihood function or conditional probability density is simply given by :

$$\mathrm{q}\left(\mathbf{X},\mathbf{Y}\right) = \frac{\exp\left(-\operatorname{tr}\left\{\mathbf{Y}\mathbf{\Phi}_{\mathsf{Y}}^{-1}\mathbf{Y}^{\dagger}\right\}\right)}{\pi^{N\times \mathsf{N_R}} \det^{\mathsf{N_R}}\left(\mathbf{\Phi}_{\mathsf{Y}}\right)} . \tag{16}$$

We will proceed by following the steps below to derive the detection metric. Consequently, to ascertain the formulation for the covariance matrix, $\mathbf{\Phi}_{\mathsf{Y}}$, we shall invoke the subsequent *theorem*.

Theorem 1. [38, Sec. 2, Th. 2].
Let $\mathbf{A}$ be an arbitrary M by N complex matrix, and let $\mathbf{R} = \mathbf{AW}$, where $\mathbf{W} \sim \mathbb{C}\mathcal{N}(0, \mathbf{I}_M)$, meaning that $W_1, \ldots, W_M$ are independent and identically distributed with independent real and imaginary parts. In that case, $\mathbf{\Phi} = \mathbb{E}\{\mathbf{AWW}^{\dagger}\mathbf{A}^{\dagger}\} = \mathbf{AA}^{\dagger}$. Therefore, $\mathbf{R} \sim \mathbb{C}\mathcal{N}(0, \mathbf{AA}^{\dagger})$.

Stated directly, the covariance matrix can be expressed as follows:

$$\begin{aligned}
\mathbf{\Phi}_{\mathsf{Y}} &\triangleq \mathbb{E}\{\mathbf{YY}^{\dagger}\}, \\
&\triangleq \mathbb{E}\{(\mathbf{XH}+\mathbf{Z})(\mathbf{XH}+\mathbf{Z})^{\dagger}\}, \\
&\triangleq \mathbb{E}\{\mathbf{XHH}^{\dagger}\mathbf{X}^{\dagger}\} + \mathbb{E}\{\mathbf{ZZ}^{\dagger}\}, \\
&\triangleq \mathbf{XX}^{\dagger} + 2\sigma^2\mathbf{I} \triangleq \mathbf{XX}^{\dagger} + \mathrm{N}_0\mathbf{I} .
\end{aligned} \tag{17}$$

This expression for the covariance matrix is commonly encountered in the literature [5], [6].

Definition 2. Let's now introduce some relevant mathematical properties, which we shall consider in the metric derivation steps.

1) $\det\left(\mathbf{I}+\mathbf{AB}\right) = \det\left(\mathbf{I}+\mathbf{BA}\right)$ [39],
2) $\det\left(\mathbf{I}+\rho\mathbf{A}\right) \approx 1 + \operatorname{tr}\left\{\rho\mathbf{A}\right\}$,
3) $\left(\mathbf{AB}\right)^{\dagger} = \mathbf{B}^{\dagger}\mathbf{A}^{\dagger}, \quad \left(\mathbf{A}^{\dagger}\right)^{\dagger} = \mathbf{A}$,
4) $\operatorname{tr}\left\{\mathbf{AB}\right\} = \operatorname{tr}\left\{\mathbf{BA}\right\}, \quad \operatorname{tr}\left\{\mathbf{A}+\mathbf{B}\right\} = \operatorname{tr}\left\{\mathbf{A}\right\} + \operatorname{tr}\left\{\mathbf{B}\right\}$,
5) $\operatorname{tr}\left\{\mathbf{A}^{-1}\mathbf{BB}^{\dagger}\right\} = \operatorname{tr}\left\{\mathbf{B}^{\dagger}\mathbf{A}^{-1}\mathbf{B}\right\}$.

Next, the determinant of $\mathbf{\Phi}_{\mathsf{Y}}$ is shown to be:

$$\det \mathbf{\Phi}_{\mathsf{Y}} = \det\left(\mathrm{N}_0\mathbf{I} + \mathbf{XX}^{\dagger}\right) . \tag{18}$$

Furthermore, the covariance matrix $\mathbf{\Phi}$ involves the addition of two matrices, which is amenable to consider the use of matrix inversion lemmas, Sherman-Morrison-Woodbury formula, or simply the Woodbury Matrix identity.

Lemma 2. [40, The Woodbury Matrix identity].

$$\left(\mathbf{A}+\mathbf{UCV}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U}\right)^{-1}\mathbf{VA}^{-1}, \tag{19}$$

where $\mathbf{A}$, $\mathbf{U}$, $\mathbf{C}$, and $\mathbf{V}$ are matrices with comfortable dimensions: $\mathbf{A}$ is a $n \times n$ matrix, $\mathbf{C}$ is a $k \times k$ matrix, $\mathbf{U}$ is a $n \times k$ matrix, and $\mathbf{V}$ is a $k \times n$ matrix.

The expression for the inverse of the covariance matrix is as follows: $\mathbf{\Phi}_{\mathsf{Y}}^{-1} = \left(\mathrm{N}_0\mathbf{I} + \mathbf{XX}^{\dagger}\right)^{-1}$.

Saying $\mathbf{A} = \mathrm{N}_0\mathbf{I}, \quad \mathbf{C} = \mathbf{I}, \quad \mathbf{U} = \mathbf{X}, \quad \mathbf{V} = \mathbf{X}^{\dagger}$, then,

$$\begin{aligned}
\mathbf{\Phi}^{-1} &= \left(\mathbf{A}+\mathbf{UCV}\right)^{-1}, \\
&= \mathrm{N}_0^{-1}\mathbf{I} - \mathrm{N}_0^{-1}\mathbf{X}\left[\mathrm{N}_0\mathbf{I} + \mathbf{X}^{\dagger}\mathbf{X}\right]^{-1}\mathbf{X}^{\dagger}, \\
&= \mathrm{N}_0^{-1}\mathbf{I} - \mathrm{N}_0^{-1}\mathbf{XDX}^{\dagger}, \quad \text{where} \quad \mathbf{D} = \left[\mathrm{N}_0\mathbf{I} + \mathbf{X}^{\dagger}\mathbf{X}\right]^{-1} .
\end{aligned} \tag{20}$$

Proposition 3. The proposed likelihood function $\mathrm{q}\left(\mathbf{X},\mathbf{Y}\right)$ can be stated as follows:

$$\begin{aligned}
\mathrm{p}\left(\mathbf{Y}|\mathbf{X}\right) &= \frac{1}{\mathbf{L}_{\mathsf{X}}}\exp\left(-\operatorname{tr}\left\{\mathbf{Y}^{\dagger}\mathbf{\Phi}_{\mathsf{Y}}^{-1}\mathbf{Y}\right\}\right) \\
&= \frac{1}{\mathbf{L}_{\mathsf{X}}}\exp\left(-\operatorname{tr}\left\{\mathbf{Y}^{\dagger}\left(\frac{1}{\mathrm{N}_0}\mathbf{I} - \frac{1}{\mathrm{N}_0}\mathbf{XDX}^{\dagger}\right)\mathbf{Y}\right\}\right),
\end{aligned} \tag{21}$$

where $\mathbf{L}_{\mathsf{X}} = \pi^{N\times \mathsf{N_R}} \det^{\mathsf{N_R}}\left(\mathrm{N}_0\mathbf{I} + \mathbf{XX}^{\dagger}\right)$.
Ignoring the multiplicative terms independent of $\mathbf{X}$, (21) reduces to:

$$\mathrm{q}\left(\mathbf{X},\mathbf{Y}\right) \approx \frac{1}{\mathbf{L}_{\mathsf{X}}}\exp\left(\frac{1}{\mathrm{N}_0}\operatorname{tr}\left\{\left(\mathbf{X}^{\dagger}\mathbf{Y}\right)^{\dagger}\mathbf{D}\left(\mathbf{X}^{\dagger}\mathbf{Y}\right)\right\}\right). \tag{22}$$

Corollary 3. As described in the proposition 2, we can incorporate the channel estimate into the metric to take the full merit of the JED principle.

For this purpose, we simply rewrite $\mathbf{X} = \mathbf{X}^{(\mathsf{d})} + \mathbf{X}^{(\mathsf{p})}$. Then, we can reveal $\widehat{\mathbf{H}}_{\mathrm{LS}}$ in the metrics:

$$\mathbf{X}^{\dagger}\mathbf{Y} = \underbrace{\mathbf{X}^{(\mathsf{p})^{\dagger}}\mathbf{Y}^{(\mathsf{p})}}_{\text{channel estimate}} + \mathbf{X}^{(\mathsf{d})^{\dagger}}\mathbf{Y}^{(\mathsf{d})} = \mathbf{C}_{\mathsf{p}}\widehat{\mathbf{H}}_{\mathrm{LS}} + \mathbf{X}^{(\mathsf{d})^{\dagger}}\mathbf{Y}^{(\mathsf{d})} , \tag{23}$$

where $\mathbf{C}_{\mathsf{p}} = \mathbf{X}^{(\mathsf{p})^{\dagger}}\mathbf{X}^{(\mathsf{p})}$ given that $\widehat{\mathbf{H}}_{\mathrm{LS}} = \frac{\mathbf{X}^{(\mathsf{p})^{\dagger}}\mathbf{Y}^{(\mathsf{p})}}{\mathbf{X}^{(\mathsf{p})^{\dagger}}\mathbf{X}^{(\mathsf{p})}}$. This channel estimate is obtained via a joint least-squares (LS) channel estimation using averaging or smoothing over the number of dimensions exhibiting channel coherence.

The joint least squares method enables the consideration of spatial correlation between the receiving and transmitting antennas, resulting in a more accurate estimation of the channel.

Consistent with this proposition, the introduced likelihood function for advanced joint estimation and detection can be subsequently formulated as follows.

$$\Lambda^{j}\left(\mathbf{Y}\right)=\max_{\mathbf{X}\,\in\,\chi_0^j}\left(\frac{1}{\mathrm{N}_0}\,\mathrm{tr}\left\{\left(\mathbf{C}_{\mathsf{p}}\widehat{\mathbf{H}}_{\mathrm{LS}}+\mathbf{X}^{(\mathsf{d})\,\dagger}\mathbf{Y}^{(\mathsf{d})}\right)^{\dagger}\mathbf{D}\left(\mathbf{C}_{\mathsf{p}}\widehat{\mathbf{H}}_{\mathrm{LS}}+\mathbf{X}^{(\mathsf{d})\,\dagger}\mathbf{Y}^{(\mathsf{d})}\right)\right\}\right)-\sum_{\mathbf{X}\,\in\,\chi_0^j}\log\left(\mathbf{L}_{\mathsf{X}}\right)$$
$$-\max_{\mathbf{X}\,\in\,\chi_1^j}\left(\frac{1}{\mathrm{N}_0}\,\mathrm{tr}\left\{\left(\mathbf{C}_{\mathsf{p}}\widehat{\mathbf{H}}_{\mathrm{LS}}+\mathbf{X}^{(\mathsf{d})\,\dagger}\mathbf{Y}^{(\mathsf{d})}\right)^{\dagger}\mathbf{D}\left(\mathbf{C}_{\mathsf{p}}\widehat{\mathbf{H}}_{\mathrm{LS}}+\mathbf{X}^{(\mathsf{d})\,\dagger}\mathbf{Y}^{(\mathsf{d})}\right)\right\}\right)+\sum_{\mathbf{X}\,\in\,\chi_1^j}\log\left(\mathbf{L}_{\mathsf{X}}\right). \tag{27}$$

$$\mathrm{q}(\mathbf{X},\mathbf{Y})=\frac{1}{\mathbf{L}_{\mathsf{X}}}\exp\left(\frac{1}{N_0}\,\mathrm{tr}\left\{\left(\mathbf{C}_{\mathsf{p}}\widehat{\mathbf{H}}_{\mathrm{LS}}+\mathbf{X}^{(\mathsf{d})\,\dagger}\mathbf{Y}^{(\mathsf{d})}\right)^{\dagger}\mathbf{D}\left(\mathbf{C}_{\mathsf{p}}\widehat{\mathbf{H}}_{\mathrm{LS}}+\mathbf{X}^{(\mathsf{d})\,\dagger}\mathbf{Y}^{(\mathsf{d})}\right)\right\}\right). \tag{24}$$

Then, the likelihood of the coded bit e_j s.t $b\in\{0,\ 1\}$ is given by

$$\mathrm{q}\left(e_j(\mathbf{X})=b,\mathbf{Y}\right)=\sum_{\mathbf{X}\,\in\,\chi_b^j}\mathrm{q}\left(\mathbf{X},\mathbf{Y}\right). \tag{25}$$

The LLR bit metric for the $j-th$ bit in BICM receiver is

$$\Lambda^{j}\left(\mathbf{Y}\right)=\log\frac{\mathrm{q}\left(e_j(\mathbf{X})=0,\mathbf{Y}\right)}{\mathrm{q}\left(e_j(\mathbf{X})=1,\mathbf{Y}\right)}. \tag{26}$$

To ease the process of implementing such a LLR bit metric in (25), one may use its *max-log approximation* version given in (27).

Furthermore, the computational complexity of the LLR metric in a BICM MIMO system is typically on the order of $\mathcal{O}\left(\mathsf{N_T}\times\mathsf{N_R}\times N_d\times\log_2\mathcal{M}\right)$. This complexity exhibits linearity with respect to the length of received data symbols (N_d), the number of transmitting ($\mathsf{N_T}$), and receiving ($\mathsf{N_R}$) antennas, as well as the size of the modulation alphabet ($\mathcal{M}$). Considering $\mathsf{N_T}=1$, we revert to the SIMO scenario, which is similar to the proposed metric in (7) for the general non-coherent fading channel by setting the relative magnitude of the LOS component $\alpha=0$, which brings the metric back to the pure Rayleigh fading case.

C. A ($\mathsf{N_R}\times 2$) MIMO within Line-Of-Sight Channel

For the sake of simplicity, let's consider a $(\mathsf{N_R}\times 2)$ MIMO configuration due to the complexity of deriving MIMO metrics within spatial dimensions $\mathsf{N_T}>2$ under LOS channel conditions with unknown phases.

Assuming a LOS channel scenario with coherence blocks covering N symbols and no antenna correlation with the channel matrix $\mathbf{H}$, the relationship between receiver and transmitter is considered to be as follows:

$$\mathbf{y}_i=\mathrm{h}_{i,1}\mathbf{x}_1+\mathrm{h}_{i,2}\mathbf{x}_2+\mathbf{z}_i,\quad i=1,2\ldots,\mathsf{N_R}. \tag{28}$$

where $\mathbf{y}_i\sim\mathbb{C}^{N\times i}$, $\{\mathbf{x}_1,\mathbf{x}_2\}\sim\mathbb{C}^{N\times 1}$ and $\mathbf{H}\sim\mathbb{C}^{\mathsf{N_R}\times 2}$.

Explicitly, the receiver signal is modelled as

$$\mathbf{y}_i=e^{j\theta_{i,1}}\mathbf{x}_1+e^{j\theta_{i,2}}\mathbf{x}_2+\mathbf{z}_i,\quad i=1,2,\ldots,N_{\mathsf{R}}. \tag{29}$$

Applying the same approach as in the preceding sections, we establish in what follows the conditional probability density function in order to determine the BICM metric corresponding to this typical MIMO LOS scenario. It should be recalled that $\theta_{i,1}$ and $\theta_{i,2}$ are unknown to the receiver and are assumed to be i.i.d. uniform random variables on $[0,2\pi)$. Thus, neglecting multiplicative terms independent of the transmitted message, the likelihood function is shown to be:

$$\mathrm{q}\left(\{\mathbf{x}_1,\mathbf{x}_2\},\mathbf{y}_i\right)\propto\int_{\theta_{i,1}}\int_{\theta_{i,2}}\exp\left(-\frac{1}{\mathrm{N}_0}\|\mathbf{y}_i-e^{j\theta_{i,1}}\mathbf{x}_1-e^{j\theta_{i,2}}\mathbf{x}_2\|^2\right)\mathrm{d}\theta_{i,2}\ \mathrm{d}\theta_{i,1}. \tag{30}$$

By expanding the ℓ^2-norms term constituting the expression of the conditional density probability, saying $\mathbf{x}_1\mathbf{x}_2^{\dagger}=\left|\mathbf{x}_1\mathbf{x}_2^{\dagger}\right|e^{j\angle\mathbf{x}_1\mathbf{x}_2^{\dagger}}$, $\mathbf{x}_1^{\dagger}\mathbf{y}_i=\left|\mathbf{x}_1^{\dagger}\mathbf{y}_i\right|e^{j\angle\mathbf{x}_1^{\dagger}\mathbf{y}_i}$, $\mathbf{x}_2^{\dagger}\mathbf{y}_i=\left|\mathbf{x}_2^{\dagger}\mathbf{y}_i\right|e^{j\angle\mathbf{x}_2^{\dagger}\mathbf{y}_i}$, and subsequently disregarding the independent terms of $\mathbf{x}_1$ and $\mathbf{x}_2$, this lends to :

$$\|\cdot\|^2\approx\|\mathbf{x}_1\|^2+\|\mathbf{x}_2\|^2+2|\mathbf{x}_1\mathbf{x}_2^{\dagger}|\cos\left(\theta_{i,1}-\theta_{i,2}+\angle\mathbf{x}_1\mathbf{x}_2^{\dagger}\right)-2|\mathbf{x}_1^{\dagger}\mathbf{y}_i|\cos\left(\theta_{i,1}-\angle\mathbf{x}_1^{\dagger}\mathbf{y}_i\right)-2|\mathbf{x}_2^{\dagger}\mathbf{y}_i|\cos\left(\theta_{i,2}-\angle\mathbf{x}_2^{\dagger}\mathbf{y}_i\right). \tag{31}$$

For reasons of simplicity, an assumption of orthogonality between the modulated symbols $\mathbf{x}_1$ and $\mathbf{x}_2$ is necessary. Actually, in MIMO systems, it is practicable or desired that the modulated symbols of distinguish antennas be orthogonal to each other. Thus, assuming orthogonality between $\mathbf{x}_1$ and $\mathbf{x}_2$, this means that $\langle\mathbf{x}_1,\mathbf{x}_2\rangle=\mathbf{0}$. Therefore, we can proceed with successive integration with respect to $\theta_{i,1}$ and $\theta_{i,2}$ using the *Fubini's Theorem* [41].

$$\mathrm{q}\left(\{\mathbf{x}_1,\mathbf{x}_2\},\mathbf{y}_i\right)\propto\exp\left(-\tfrac{\|\mathbf{x}_1\|^2+\|\mathbf{x}_2\|^2}{\mathrm{N}_0}\right)\int_{\theta_{i,1}}\exp\left(\frac{2}{\mathrm{N}_0}\left|\mathbf{x}_1^{\dagger}\mathbf{y}_i\right|\cos\left(\theta_{i,1}-\angle\mathbf{x}_1^{\dagger}\mathbf{y}_i\right)\right)\int_{\theta_{i,2}}\exp\left(\frac{2}{\mathrm{N}_0}\left|\mathbf{x}_2^{\dagger}\mathbf{y}_i\right|\cos\left(\theta_{i,2}-\angle\mathbf{x}_2^{\dagger}\mathbf{y}_i\right)\right)\mathrm{d}\theta_{i,2}\ \mathrm{d}\theta_{i,1}.$$

Proposition 4. The likelihood function is shown to be

$$\mathrm{q}\left(\{\mathbf{x}_1,\mathbf{x}_2\},\mathbf{y}\right)\propto\prod_{i=0}^{\mathsf{N_R}-1}\exp\left(-\frac{\|\mathbf{x}_1\|^2+\|\mathbf{x}_2\|^2}{\mathrm{N}_0}\right)\times\mathrm{I}_0\left(\frac{2}{\mathrm{N}_0}|\mathbf{x}_1^{\dagger}\mathbf{y}_i|\right)\times\mathrm{I}_0\left(\frac{2}{\mathrm{N}_0}|\mathbf{x}_2^{\dagger}\mathbf{y}_i|\right). \tag{32}$$

The likelihood of the coded bit $e_j\in\{0,1\}$ is

$$\mathrm{q}\left(e_j(\{\mathbf{x}_1,\mathbf{x}_2\})=b,\mathbf{y}_i\right)=\sum_{\{\mathbf{x}_1,\mathbf{x}_2\}\,\in\,\chi_b^j}\mathrm{q}\left(\{\mathbf{x}_1,\mathbf{x}_2\},\mathbf{y}_i\right). \tag{33}$$

Thus, the LLR bit metric for the $j-th$ coded bit is as follows

$$\Lambda^{j}\left(\mathbf{y}_i\right)=\log\frac{\mathrm{q}\left(e_j(\{\mathbf{x}_1,\mathbf{x}_2\})=0,\mathbf{y}_i\right)}{\mathrm{q}\left(e_j(\{\mathbf{x}_1,\mathbf{x}_2\})=1,\mathbf{y}_i\right)}. \tag{34}$$

$$\Lambda^{j}\left(\mathbf{y}\right) = \max_{\{\mathbf{x}_1,\mathbf{x}_2\}\,\in\,\chi_0^j} \sum_{i=0}^{\mathsf{N_R}-1} \frac{2}{\mathrm{N}_0}\left(\left|N_p\hat{\mathrm{h}}_{i,1}^{\mathrm{LS}} + \ \mathbf{x}_1^{(\mathsf{d})\dagger}\mathbf{y}_i^{(\mathsf{d})}\right| + \left|N_p\hat{\mathrm{h}}_{i,2}^{\mathrm{LS}} + \ \mathbf{x}_2^{(\mathsf{d})\dagger}\mathbf{y}_i^{(\mathsf{d})}\right| - \frac{\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2}{2}\right) - \max_{\{\mathbf{x}_1,\mathbf{x}_2\}\,\in\,\chi_1^j} \sum_{i=0}^{\mathsf{N_R}-1} \frac{2}{\mathrm{N}_0}\left(\left|N_p\hat{\mathrm{h}}_{i,1}^{\mathrm{LS}} + \ \mathbf{x}_1^{(\mathsf{d})\dagger}\mathbf{y}_i^{(\mathsf{d})}\right| + \left|N_p\hat{\mathrm{h}}_{i,2}^{\mathrm{LS}} + \ \mathbf{x}_2^{(\mathsf{d})\dagger}\mathbf{y}_i^{(\mathsf{d})}\right| - \frac{\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2}{2}\right). \tag{35}$$

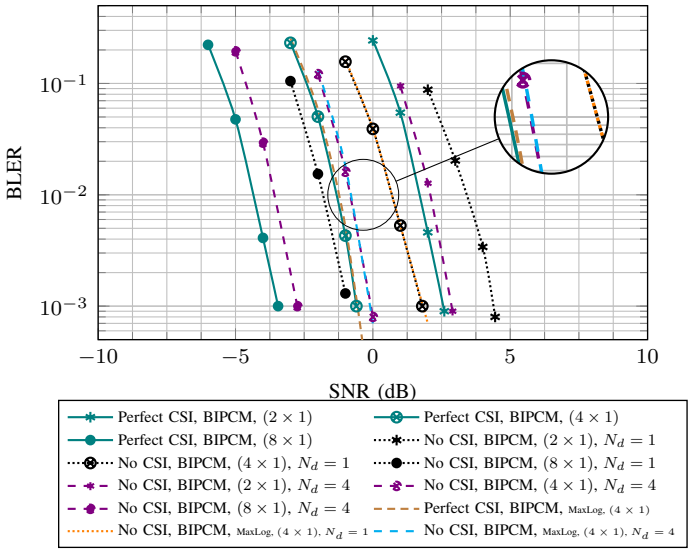


Fig. 4. Block Error Rate, 48 bits(TBs+CRC), NR POLAR BICM, R=48/64, (CRC)-aided successive-cancellation list decoder(List length=8), QPSK modulation, 1 OFDM symbol, 4 PRBs, 48 REs (32 data, 16 DMRS), $([8,4,2]\times 1)$ SIMO, vs outer (MC) and inner (RCUs) bounds, $\alpha = 1$, Line-of-Sight (LOS) channel.

For simpler processing, the *max-log approximation* of (34) is provided in (35).

Remark 3. The computational complexity of the LLR metric in such a BICM system is typically on the order of $\mathcal{O}\left(2 \times \mathsf{N_R} \times N_d \times \log_2 \mathcal{M}\right)$. Considering $\mathsf{N_T} = 1$, we revert to the SIMO scenario, which is similar to the metric we presented in (8) for the general non-coherent fading channel by setting the relative magnitude of the LOS component $\alpha = 1$, which is amenable to the pure LOS scenario. Furthermore, if the orthogonality between $\mathbf{x}_1, \mathbf{x}_2$ is not assumed, it would be quite challenging to establish a metric for such a LOS channel when $\theta_{i,j}$ is unknown.

IV. Numerical Results

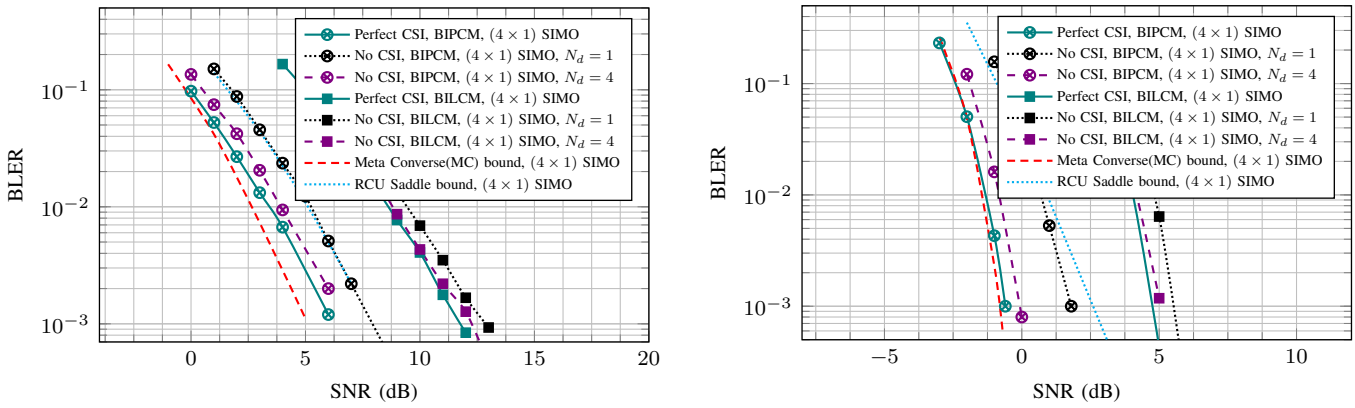
A. Performance Analysis

For illustrative purposes, we consider three distinct configurations: (4×1) SIMO BICM, (2×2) and (4×2) MIMO BICM. The subsequent figures show the performance of BIPCM/BILCM with joint estimation and detection under various channel conditions. The evaluation discerns performance disparities across three scenarios: *Perfect CSI*, *No CSI* $(N_d = 4)$, and *No CSI* $(N_d = 1)$. The simulations employ NR POLAR and NR LDPC coding schemes, coupled with QPSK modulation. The transmission process encompasses a transport block length of 48 bits.The resource allocation procedure employs a singular OFDM symbol with either 4 PRBs or 48 resource elements (comprising 32 REs for data components and 16 REs for DMRS components). The DMRS sequences occupy 4 REs per PRB. This transmission framework, featuring joint transmission of reference and data signals within shared OFDM symbols, adheres to the standard practice in *physical uplink control channel*, *physical uplink shared channel*, as well as in some downlink control channels. Hence, from the perspective of 5G NR PHY layers, the transmission employing BIPCM aligns with a PUCCH format 2 transmission, while that employing BILCM corresponds to a PUSCH transmission.

The results illustrated in Figure 4 delineate the performance of the $([8,4,2]\times 1)$ SIMO BICM systems, with joint estimation and detection, over a LOS channel. This evaluation is conducted with a focus on the scenario where $\alpha = 1$, aiming to discern the performance disparity between the *Perfect CSI* and *No CSI* conditions, particularly in coverage scenarios marked by low signal-to-noise ratios.

Note that the *No CSI*$(N_d = 1)$ case also refers to the conventional receiver and *No CSI JED*$(N_d = 4)$ corresponds to the proposed receiver based on *joint estimation and detection*$(N_d = 4)$. The JED-based receiver yields a performance gain of 1.25 dB, 1.5 dB and 1.75 dB with respect to the conventional receiver over 2, 4 and 8 receive antennas respectively at a BLER threshold of 1%. From this insight, it is apparent that when the number of antennas increases, the performance gap between the *Perfect CSI* and the *No CSI* situations (e.g., JED-based receiver) expands. Additionally, the max-log metric performs nearly as well as the accurate metric(i.e, Log-based LLRs) since Gray-mapped constellations are in use. Thus, the max-log metric seems to have minimal impact on receiver performance, as we operate with low modulation orders.

In addition, to conduct a comprehensive comparative analysis of our findings, we have employed finite block length bounds, integrating both converses and achievability results as established in the scientific literature [16]–[20]. For a more insight into the *metaconverse* (MC) and *random coding union* (RCU) bounds used in the aforementioned figures, reference can be made to the following scholarly contributions [16], [19], [20]. Figure 5a shows that using the joint estimation and detection (JED) approach enhances performance by 1.5 dB and 0.75 dB for polar and LDPC coded configurations in a (4×1) SIMO system with Rayleigh block fading channel. Additionally, a performance difference of approximately 1.25 dB is observed between our JED-based receiver $(N_d = 4)$ utilizing BIPCM and the metaconverse bound. Interestingly,



(a) Rayleigh Block Fading channel.

(b) Line-of-Sight (LOS) channel.

Fig. 5. Block Error Rate, 48 bits(TBs+CRC), NR POLAR BICM (CRC-aided successive-cancellation list decoder, List length=8), NR LDPC BICM (belief propagation decoder, iteration=30) QPSK modulation, 1 OFDM symbol, 4 PRBs, 48 REs (32 data, 16 DMRS), (4×1) SIMO, vs outer (MC) and inner (RCUs) bounds.

this difference diminishes to only 0.5 dB when compared to the performance of the coherent receiver with *Perfect CSI*. Conversely, referring to Figure 5b, which illustrates (4×1) BIPCM SIMO and BILCM SIMO configurations over a LOS channel scenario. In the first configuration, at a 1% BLER, the performance gap between the metaconverse bound and the JED-based receiver $(N_d = 4)$ is 0.7 dB. This stands in contrast to a 2.2 dB difference observed within the *No-CSI*$(N_d = 1)$. Despite having similar code rates transmission parameters, BIPCM consistently outperforms BILCM. This is attributed, in part, to the optimization of the 3GPP polar code for short block lengths, whereas the 3GPP LDPC code is designed for longer transport block lengths.

Furthermore, the results illustrated in Figure 6a, representative of a (2×2) spatially multiplexed MIMO configuration, align with the trends observed in the preceding figures. Accordingly, the notable performance enhancements are significant, indicating improvements of 0.6 dB and 0.3 dB with JED-based receivers $(N_d = 4)$ when employing BIPCM and BILCM, respectively.

Likewise, figure 6b illustrates the (4×2) BIPCM and BILCM MIMO configurations over LOS channel. At 1% of BLER, the performance comparison in the first configuration indicates a gain of 1 dB with the JED-based receiver over the conventional receiver. Hence, there is a gap of 0.6 dB between the JED-based receiver $(N_d = 4)$ and the *Perfect CSI*-based receiver. In the second configuration, namely the (4×2) BILCM MIMO, there is a 0.75 dB improvement with the JED-based receiver $(N_d = 4)$ compared to the conventional receiver, and a 0.5 dB gap between the *Perfect CSI*-based receiver and the JED-based receiver $(N_d = 4)$.

Remarkably, it can be contended that the advanced receiver outperforms the conventional counterpart and demonstrates greater resilience in the face of inaccurate channel estimation.

B. Metric Performance Extra Enhancement

1) DMRS Power Boosting

DMRS power boosting was extensively discussed in our prior correspondence [11], particularly in scenarios where reference and data symbols are jointly conveyed in common OFDM symbols. Conceptually, envision the signal as comprising a data component and a data-independent component, or pilots, in a frequency-interleaved fashion.

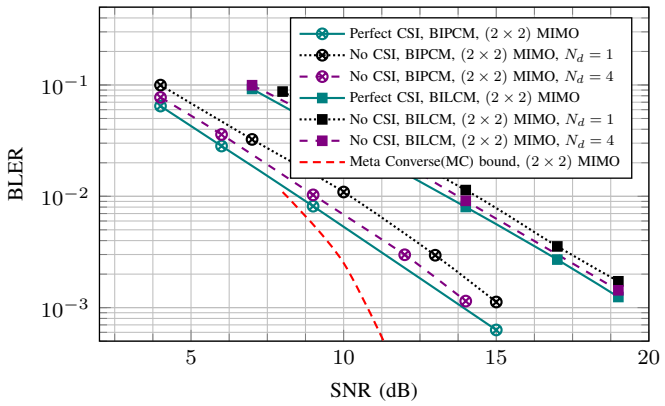
To enhance the power of pilot signals within an interleaved set, scaling the power of DMRSs while keeping the data signals unchanged or constant is crucial. Put simply, the boosted transmitted signal, denoted as $\mathbf{x}_{\texttt{boosted}}$, is then defined as $\mathbf{x}_{\texttt{boosted}} = \mathbf{x}^{(\mathsf{d})} + \beta\ \mathbf{x}^{(\mathsf{p})}$. The adaptive power adjustment procedure is contingent on β values and aims to increase the power or strength of the pilot signals within the composite signal. Care should be taken to select an appropriate value for β to achieve the desired power augmentation without introducing distortion or signal saturation. To comply with potential radio frequency constraints, β must be perfectly calibrated. Optimal performance enhancement is achieved when β is set to 1.75 (representing a 75% increase in DMRS power compared to its initial value). In Figure 7a, it is apparent to see that the proposed JED-based receiver yields an additional gain of approximately 0.5 dB.

This improvement consistently converges towards an ever-closer alignment with the performance benchmarks set by the ideal or coherent receiver, mainly through adaptive DMRS/-data power adjustment. In this spirit, we can even prioritise transmission with fewer DMRS in order to *bootstrap* the advance receiver, therefore reducing some transmission overhead.

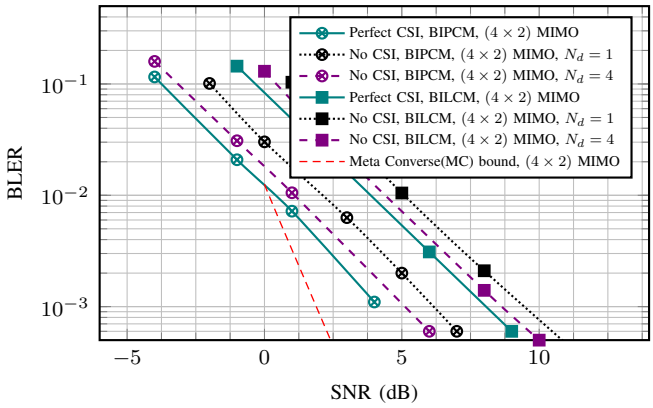
The implications of slightly adjusting the DMRS power within the 3GPP standard are significant. Specifically, it is feasible to allow the *user equipment* (UE) to adjust the power allocation between the DMRS and data transmission. This flexibility in adaptive DMRS power adjustment is somewhat transparent to the receiver.

2) Iterative Advanced Joint Estimation Detection

Iterative channel estimation detection is extensively investigated in the literature under various facets, showing the merits of the iterative process in achieving progressive performance. Considering recent advances in this field, Jiao *et al.* [42] introduce a joint channel estimation and decoding scheme for polar coded *sparse code multiple access* (SCMA) systems over fading channels in the wake of enhancing 5G communications for *massive machine-type communications* (mMTC) within



(a) (2×2) SU-MIMO, vs outer (MC) bound, Rayleigh Block Fading channel.



(b) (4×2) SU-MIMO, vs outer (MC) bound, Line-of-Sight (LOS) channel.

Fig. 6. Block Error Rate, 48 bits(TBs+CRC), NR POLAR BICM (CRC-aided successive-cancellation list decoder, List length=8), NR LDPC BICM (belief propagation decoder, iteration=30) QPSK modulation, 1 OFDM symbol, 4 PRBs, 48 REs (32 data, 16 DMRS).

the *Internet of Things* (IoT). Their proposed joint estimation-decoding scheme relies on *sparse Bayesian learning* (SBL) for initial CSI measurements, iteratively refines channel estimation and detection for enhanced system performance. Hence, the iterative joint estimation and detection procedure aiming to be put forward in this work aligns with existing literature, underscoring the importance of iterative strategies in both channel estimation and detection stages. Particularly, within BICM receiver framework, the iterative process is employed initially to enhance channel estimation quality and subsequently applied to the soft-detector to achieve superior and high-quality LLRs. The iterative process may spread over two layers.Merely through the iterative channel estimation process, one can achieve a gain of few dBs. The iterative procedure, when applied to an advanced detector leading to a fully iterative receiver, further enhances these gains by a few additional dBs. Nevertheless, it is noteworthy that the proposed JED-based receiver without iteration and with iteration exhibit a considerable degree of similarity in terms of performance, as depicted in Figure 7b. Specifically, there is an approximate difference of 0.15 dB when considering 10 iterations. This observation underscores the robustness of the proposed JED-based receiver in terms of channel estimation and detection, even in the absence of an iterative process. Indeed, the estimated channel and the employed detection strategy appear to be robust to a certain extent. In some scenarios, relying solely on the advanced JED without introducing an iterative process may be sufficient. This approach offers added value in terms of reducing receiver complexity, as introducing excessive operations at the receiver level could incur additional overhead and latency.

3) Performance Analysis within Various Modulation Sets We evaluate performance on various M-ary PSK and M-ary QAM modulations, namely BPSK, QPSK, 8-PSK and 16-QAM. In the different coded modulation situations, we need to adapt the transmission parameters, in particular the number of PRBs or resource elements required. Indeed, considering the same coding rate R, at the output of the modulator the modulated symbols from the different scenarios have not the same length. Table I, gives an inventory of the transmission parameters according to different modulation configurations, determining the number of resource elements for data and DMRS that will be required.

TABLE I
INVENTORY OF THE TRANSMISSION PARAMETERS WITHIN M-ARY PSK/QAM MODULATION SETS

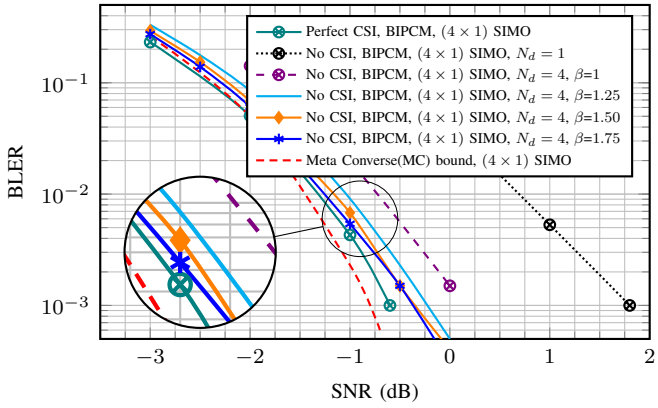
Modulations	Rate	REs: N_d	REs: N_p,	N_{PRB}
BPSK	48/96	96	48	12
QPSK	48/96	48	24	6
8-PSK	48/96	32	16	4
16-QAM	48/96	24	12	3

C. Complexity Analysis

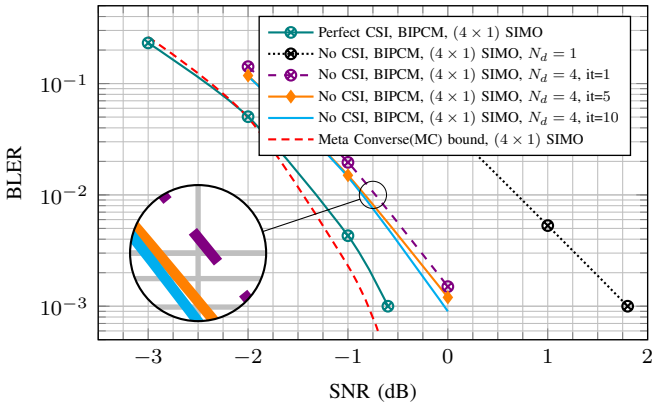
The complexity of the detection metrics is analysed using Monte Carlo simulation. The execution time highlights the time elapsed between the input and output of the demodulator, concisely, until the LLRs are generated. It is relevant to ascertain the block size range wherein complexity is comparatively diminished when contrasted with traditional metrics to establish a better trade-off between performance and complexity. Analytically, within a very short block regime, the standard and advanced receiver metrics exhibit near equivalence, as illustrated in Figure 9a. However, as the block size increases, the complexity of the advanced receivers becomes greater than that of the conventional receivers, hence the need for block detection to break down this increasing complexity. By harnessing parallel processing and assuming independence among the (N_d)-dimensional blocks making up the detection windows, the detection complexity of the JED oriented receivers can be reduced. This reduction is directly correlated with the degree of parallelism, which is contingent not only upon the quantity of processing nodes but also on the efficient distribution of the detection window processing among these nodes. Consequently, this approach unquestionably accelerates the overall detection process at the receiver level.

V. CONCLUSIONS

This paper presented novel bit-interleaved coded modulation metrics for joint estimation detection using a training or reference signal transmission strategy for short block length channels. We showed that it is possible to enhance the performance and sensitivity of advanced receivers, especially when channel state information is unknown and the density



(a) DMRS power Boosting via a scaling factor β.



(b) Iterative estimation-detection spanning from 5 and 10 iterations.

Fig. 7. Block Error Rate, 48 bits(TBs+CRC), NR POLAR BICM, (CRC)-aided successive-cancellation list decoder, QPSK modulation, 1 OFDM symbol, 4 PRBs, 48 REs, (40REs= data, 8REs= DMRS), proposed JED-based receiver($N_d = 4$) over (4 × 1) SIMO, on Line-of-Sight (LOS) Channel.

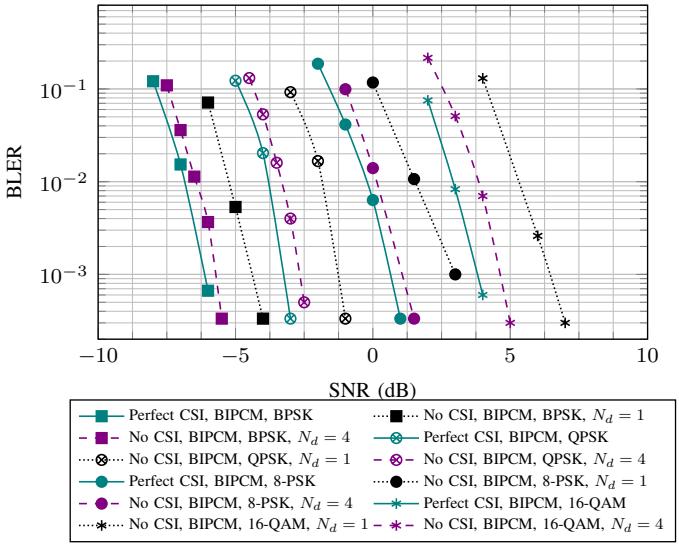


Fig. 8. Block Error Rate, 48 bits(TBs+CRC), NR POLAR BICM, R=48/64, (CRC)-aided successive-cancellation list decoder(List length=8), 1 OFDM symbol, BPSK, QPSK, 8-PSK modulations,(4 × 1) SIMO, $\alpha = 1$, Line-of-Sight (LOS) channel.

of training dimensions is low. The proposed techniques take advantage of joint estimation/detection. The performance analysis made use of a full 5G transmitter and receiver chain for both Polar and LDPC coded transmissions paired with M-ary PSK modulation schemes. We considered transmissions where reference signals are interleaved with data and both are transmitted over a small number of OFDM symbols so that near-perfect channel estimation cannot be achieved. This is particularly adapted to mini-slot transmissions for ultra-reliable low-latency communications or short-packet random-access use-cases. We characterized the performance for up to SIMO and MIMO configurations in order to determine the performance gain offered by the proposed BICM detection in realistic base station receiver scenarios. Our findings demonstrated that when the detection windows used in the metric units is on the order of four modulated symbols the proposed BICM metrics can be used to achieve detection performance that is close to that of a coherent receiver with perfect channel state information for both polar and LDPC coded configurations.

APPENDIX

A. Metric derivation for general non-coherent fading channel

Assuming that θ_i is unknown and randomly distributed over $[0, 2\pi)$, the conditional probability density function is defined as

$$\mathrm{q}\left(\mathbf{x}, \mathbf{y}_i\right) = \frac{1}{2\pi \det \Phi} \int_0^{2\pi} \exp\left(-\frac{1}{2}\left(\mathbf{y}_i - \mu\{\mathbf{x}, \theta_i\}\right)^{\dagger} \Phi^{-1}\left(\mathbf{y}_i - \mu\{\mathbf{x}, \theta_i\}\right)\right) \mathrm{d}\theta_i \ . \tag{36}$$

Saying $\mu\{\mathbf{x}, \theta_i\} = \sqrt{\alpha} e^{j\theta_i}\mathbf{x}$, then

$$\mathrm{q}\left(\mathbf{x}, \mathbf{y}_i\right) = \frac{1}{2\pi \det \Phi} \int_0^{2\pi} \exp\left(-\frac{1}{2}\left(\mathbf{y}_i - \sqrt{\alpha} e^{j\theta_i}\mathbf{x}\right)^{\dagger} \tag{37}$$

$$\Phi^{-1}\left(\mathbf{y}_i - \sqrt{\alpha} e^{j\theta_i}\mathbf{x}\right)\right) \mathrm{d}\theta_i \ . \tag{38}$$

Covariance Matrix:

Knowing that $\mathbf{y}_i - \sqrt{\alpha} e^{j\theta_i}\mathbf{x} = \sqrt{1-\alpha}\mathrm{h}_{i,f}\mathbf{x} + \mathbf{z}_i$, then

$$\begin{aligned} \Phi &\triangleq \frac{1}{2}\mathbb{E}\left[\left(\sqrt{1-\alpha}\mathrm{h}_i^{(f)}\mathbf{x} + \mathbf{z}_i\right)\left(\sqrt{1-\alpha}\mathrm{h}_i^{(f)}\mathbf{x} + \mathbf{z}_i\right)^{\dagger}\right], \\ &\triangleq (1-\alpha)\mathbf{x}\mathbf{x}^{\dagger}\sigma_h^2 + \sigma_z^2\mathbf{I}_N, \text{where } \sigma_h^2 = 1, \\ &\triangleq (1-\alpha)\mathbf{x}\mathbf{x}^{\dagger} + \frac{\mathrm{N}_0}{2}\mathbf{I}_N \ . \end{aligned} \tag{39}$$

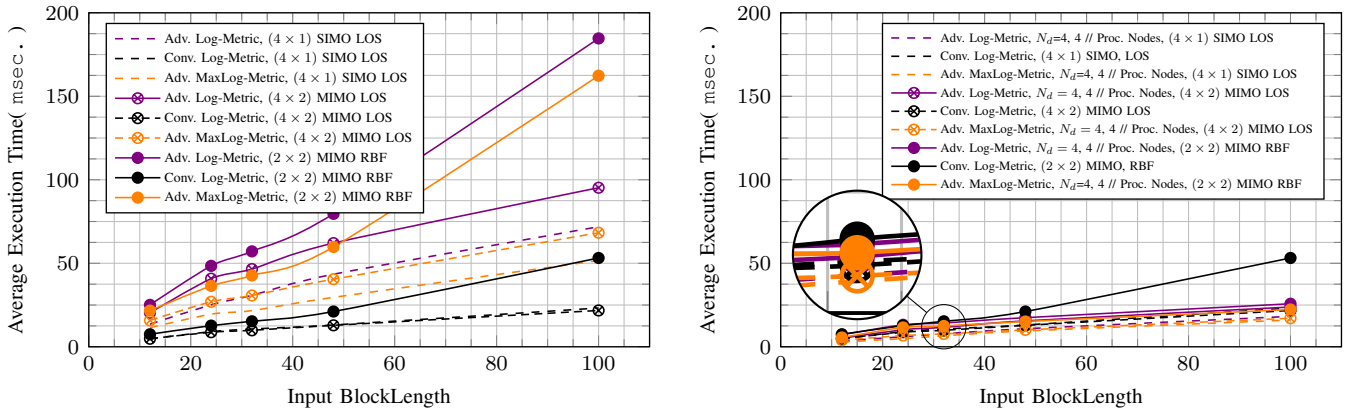
Determinant:

$$\begin{aligned} \det \Phi &= \det\left((1-\alpha)\mathbf{x}\mathbf{x}^{\dagger} + \sigma_z^2\mathbf{I}\right), \\ &= \det\left(\sigma_z^2\mathbf{I} + (1-\alpha)\mathbf{x}\mathbf{x}^{\dagger}\right), \\ &= \frac{1}{2}\left(\mathrm{N}_0 + 2(1-\alpha)\left\|\mathbf{x}\right\|^2\right) \ . \end{aligned} \tag{40}$$

Inverse of Φ :

The matrix inversion *Lemma 1* must be used. Note that, here, we have a special case where $\mathbf{V}$, $\mathbf{U}$ are vectors, consequently $\mathrm{rank}\{\mathbf{x}^{\dagger}\mathbf{x}\} = 1$.

$$\text{Saying :} \quad \mathbf{A} = \sigma_z^2\mathbf{I} \quad \mathbf{C} = (1-\alpha)\mathbf{I} \quad \mathbf{U} = \mathbf{x} \quad \mathbf{V} = \mathbf{x}^{\dagger}. \tag{41}$$



(a) Block detections spanning the detection windows(N_d=4) within sequential processing.

(b) Block detections spanning the detection windows(N_d=4) over 4 parallel processing nodes.

Fig. 9. Average time complexity of the LLR computation within 10, 000 iterations in the short block-length regime ($12-100$ bits) comparing the conventional metric with that proposed in log and max-log domain within (4×1) SIMO, (4×2) SU-MIMO on Line-of-Sight (LOS) Channel, and (2×2) SU-MIMO on Rayleigh Block Fading (RBF) channels.

$$\begin{aligned}\Phi^{-1} &= \left(\mathbf{A}+\mathbf{UCV}\right)^{-1}, \\ &= \frac{2}{\mathrm{N}_0} - \frac{2}{\mathrm{N}_0}\mathbf{x}\left(\frac{2(1-\alpha)}{\mathrm{N}_0+2(1-\alpha)\left\|\mathbf{x}\right\|^2}\right)\mathbf{x}^{\dagger}\ .\end{aligned} \tag{42}$$

Saying $\beta_x = \frac{2(1-\alpha)}{\mathrm{N}_0(\mathrm{N}_0+2(1-\alpha)\|\mathbf{x}\|^2)}$, then $\Phi^{-1} = \frac{2}{\mathrm{N}_0} - 2\mathbf{x}\beta_x\mathbf{x}^{\dagger}$.

Likelihood function :

$$\begin{aligned}&\mathrm{q}\left(\mathbf{x},\mathbf{y}_i\right) \\ &= \frac{\displaystyle\int_0^{2\pi}\exp\left(-\frac{1}{2}\left(\mathbf{y}_i-\mu\right)^{\dagger}\left(\frac{2}{\mathrm{N}_0}-2\mathbf{x}\beta_x\mathbf{x}^{\dagger}\right)\left(\mathbf{y}_i-\mu\right)\right)}{2\pi\det\Phi}\mathrm{d}\theta_i, \\ &= \frac{\displaystyle\int_0^{2\pi}\exp\left(-\frac{1}{\mathrm{N}_0}\left|\mathbf{y}_i-\mu\right|^2+\beta_x\left|\left(\mathbf{y}_i-\mu\right)^{\dagger}\mathbf{x}\right|^2\right)}{2\pi\det\Phi}\mathrm{d}\theta_i.\end{aligned} \tag{43}$$

By extending the terms into the exponential,

ignoring those that are independent of $\mathbf{x}$, the likelihood function is equivalent to

$$\begin{aligned}\mathrm{q}\left(\mathbf{x},\mathbf{y}_i\right) = \frac{1}{2\pi\det\Phi}\exp\left(-\alpha\left\|\mathbf{x}\right\|^2\left(\frac{1}{\mathrm{N}_0}-\beta_x\left\|\mathbf{x}\right\|^2\right)\right. \\ \left.\beta_x\left|\mathbf{x}^{\dagger}\mathbf{y}_i\right|^2\right)\int_0^{2\pi}\exp\left(2\sqrt{\alpha}\left(\frac{1}{\mathrm{N}_0}-\beta_x\left\|\mathbf{x}\right\|^2\right)\right. \\ \left.\left|\mathbf{x}^{\dagger}\mathbf{y}_i\right|\cos\left(\phi_i+\theta_i\right)\right)\mathrm{d}\theta_i.\end{aligned} \tag{44}$$

Knowing that $\frac{1}{\pi}\int_{\varphi=0}^{\pi}\exp(zcos(\varphi))\mathrm{d}\varphi = \mathrm{I}_0(\mathrm{z})$ [43].

Saying $\mathbf{L}_{\mathrm{x}} = \mathrm{N}_0 + 2(1-\alpha)\left\|\mathbf{x}\right\|^2$, and then after ignoring multiplicative term that are independent of $\mathbf{x}$, it comes

$$\begin{aligned}\mathrm{q}\left(\mathbf{x},\mathbf{y}_i\right) \propto \frac{1}{\mathbf{L}_{\mathrm{x}}}\exp\left(-\alpha\left\|\mathbf{x}\right\|^2\left(\frac{1}{\mathrm{N}_0}-\beta_x\left\|\mathbf{x}\right\|^2\right)\right. \\ \left.+\beta_x\left|\mathbf{x}^{\dagger}\mathbf{y}_i\right|^2\right)\times\mathrm{I}_0\left(2\sqrt{\alpha}\left(\frac{1}{\mathrm{N}_0}-\beta_x\left\|\mathbf{x}\right\|^2\right)\left|\mathbf{x}^{\dagger}\mathbf{y}_i\right|\right).\end{aligned} \tag{45}$$

Expressing β_x w.r.t $\mathbf{L}_{\mathrm{x}}$, we have the relation

$$\beta_x = \frac{1}{\left\|\mathbf{x}\right\|^2\mathrm{N}_0} - \frac{1}{\left\|\mathbf{x}\right\|^2\mathbf{L}_{\mathrm{x}}}. \tag{46}$$

B. Finit-Blocklength Bounds

Herein, we present the finite-blocklength information theory tools. An outer bound, derived from the metaconverse theorem (cf. [17, Th. 28]), is introduced, while an inner bound is established using the RCUs bound [20, Th. 1]. In accordance with an analogical framework, the following is grounded in the scholarly contributions [16], [19], [20]. We strive to closely fit these state-of-the-art converse and achievability bounds to our transmission scenario.

Theorem 3. [20, RCU bound, Th. 1].
Let use capital bold letters X, for random variables (RVs) and their lower case counterparts, e.g., x, for their realizations. let denote the random vectors via $\mathbf{X} = [\mathrm{X}_1, X_2, \ldots, \mathrm{X}_N]$, and their vector realizations via $\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_N]$.
To set the achievability bound, let define the generalized information density as

$$\imath_s\left(\mathbf{x},\mathbf{y}\right) = \ln\frac{\mathrm{q}\left(\mathbf{x},\mathbf{y}\right)^s}{\mathbb{E}\left[\mathrm{q}\left(\mathbf{X}',\mathbf{y}\right)^s\right]}. \tag{47}$$

In the domain of information theory, the Gallager exponent, represented by $s > 0$, characterizes a pivotal factor. The expectation relates to the random vector $\mathbf{X}'$ having N-dimensional i.i.d. components. Over a memoryless channel, the decoding metric $\mathrm{q}\left(\mathbf{x},\mathbf{y}\right) = \prod_{n=1}^{N}\mathrm{q}\left(\mathrm{x}_n,\mathrm{y}_n\right)$. The random coding unions (RCUs) posits that, for a specified rate R, the `upper bound` on the average error probability is defined as:

$$\epsilon \le \inf_{s>0}\mathbb{E}\left[e^{-\left[\imath_s(\mathbf{X},\mathbf{Y})-\ln\left(2^{R\times N}-1\right)\right]^+}\right], \tag{48}$$

where $(a)^+ \triangleq \max(0,a)$.
In a pilot-assisted transmission, with the N_d data symbols uniformly distributed on a shell in $\mathbb{C}^{N_d}$, the maximum likelihood decoding metric is shown to be $\mathrm{q}(\mathbf{x},\mathbf{y}) = p_{\mathbf{Y}^{(\mathrm{d})}|\mathbf{X}^{(\mathrm{d})},\hat{\mathbf{H}}}\left(\mathbf{y}^{(\mathrm{d})} \mid \mathbf{x}^{(\mathrm{d})},\hat{\mathbf{h}}\right)$. Besides, the underlying PAT decoding metric, as outlined in Östman's work [19, Sec. III.D]], is expressed as $\mathrm{q}(\mathbf{x},\mathbf{y}) \propto \exp\left(-\|\mathbf{y}^{(\mathrm{d})} - \hat{\mathbf{h}}\mathbf{x}^{(\mathrm{d})}\|^2\right)$.

Theorem 4. [16, Converse bound, Sec. IV].
To set the converse bound, consider :

$$\jmath_s(\mathbf{x},\mathbf{y}) = \ln \frac{p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})}{\mathrm{q}^s_{\mathbf{Y}}(\mathbf{y})}, \tag{49}$$

where $\mathrm{q}^s_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\mu(s)}\mathbb{E}\left[p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{X}')^s\right]^{1/s}$, and the normalization factor $\mu(s)$ is selected such that $\mathrm{q}^s_{\mathbf{Y}}(\mathbf{y})$ integrates to unity.

Subsequently, for a given rate R, the `lower bound` on the average error probability is given as follows.

$$\epsilon \geq \sup_{s>0}\sup_{\lambda>0} \mathbb{P}\left[\jmath_s(\mathbf{X},\mathbf{Y}) \leq \lambda\right] - e^{\lambda - R\times N}. \tag{50}$$

In our simulations, we do not engage in optimization over the parameter s; instead, we opt for the fixed value $s = 1$, which offers a more relaxed constraint.

REFERENCES

[1] E. Zehavi, "8-PSK trellis codes for a Rayleigh fading channel",IEEE Transactions on Communication, vol.40, pp. 873–883, 1992, May 1992.

[2] A. G. i Fàbregas, A Martinez and G. Caire, "Bit-Interleaved Coded Modulation", Foundations and Trends® in Communications and Information Theory Vol. 5: No. 1–2, pp 1-153, November 2008.

[3] G. Caire, G. Taricco and E. Biglieri, " Bit- Interleaved Coded Modulation," IEEE Transactions on Information Theory, vol. 44, pp. 927–946, May 1998.

[4] B. M Hochwald and S. T Brink, "Achieving near-capacity on a multiple-antenna channel," in IEEE Transactions on Communications, vol. 51, no. 3, pp. 389-399, March 2003.

[5] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," in IEEE Transactions on Information Theory, vol. 45, no. 1, pp. 139-157, Jan. 1999.

[6] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," in IEEE Transactions on Information Theory, vol. 46, no. 2, pp. 543-564, March 2000.

[7] D. F. Carrera, D. Zabala-Blanco, C. Vargas-Rosales and C. A. Azurdia-Meza, "Extreme Learning Machine-Based Receiver for Multi-User Massive MIMO Systems," in IEEE Communications Letters, vol. 25, no. 2, pp. 484-488, Feb. 2021.

[8] S. Cammerer, F. A. Aoudia, J. Hoydis, A. Oeldemann, A. Roessler, T. Mayer and A. Keller, "A Neural Receiver for 5G NR Multi-user MIMO," ArXiv abs/2312.02601,2023.

[9] B. Lee, S. Park, D. J. Love, H. Ji and B. Shim, "Packet Structure and Receiver Design for Low Latency Wireless Communications With Ultra-Short Packets," in IEEE Transactions on Communications, vol. 66, no. 2, pp. 796-807, Feb. 2018.

[10] M. Sy, R. Knopp, "Enhanced Low-Complexity Receiver Design for Short Block Transmission Systems," 34th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications(PIMRC 2023), Toronto, ON, Canada, Sept. 2023.

[11] M. Sy, R. Knopp, "Novel Joint Estimation and Decoding Metrics for Short-Blocklength Transmission Systems," 2023 IEEE Conference on Standards for Communications and Networking (CSCN), Munich, Germany, Nov. 2023.

[12] N. Doan, "Low-complexity decoding of short linear block codes with machine learning," PhD Thesis Dissertation, McGill University, May 2022.

[13] C. Yue, V. Miloslavskaya, M. Shirvanimoghaddam, B. Vucetic and Y. Li, "Efficient Decoders for Short Block Length Codes in 6G URLLC," IEEE Communications Magazine, Vol. 61, no.4, pp 84–90, April 2023.

[14] T. -H. Vu, T. -T. Nguyen, Q. -V. Pham, D. B. da Costa and S. Kim, "A Novel Partial Decode-and-Amplify NOMA-Inspired Relaying Protocol for Uplink Short-Packet Communications," in IEEE Wireless Communications Letters, vol. 12, no. 7, pp. 1244-1248, July 2023.

[15] Jalali and Z. Ding, "Joint Detection and Decoding of Polar Coded 5G Control Channels," in IEEE Transactions on Wireless Communications, vol. 19, no. 3, pp. 2066-2078, March 2020.

[16] M. Xhemrishi, M. C. Coşkun, G. Liva, J. Östmanl and G. Durisi, "List Decoding of Short Codes for Communication over Unknown Fading Channels," 2019 53rd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2019, pp. 810-814.

[17] Y. Polyanskiy, H. V. Poor and S. Verdu, "Channel Coding Rate in the Finite Blocklength Regime," in IEEE Transactions on Information Theory, vol. 56, no. 5, pp. 2307-2359, May 2010.

[18] G. Durisi, T. Koch, J. Östman, Y. Polyanskiy and W. Yang, "Short-Packet Communications Over Multiple-Antenna Rayleigh-Fading Channels," in IEEE Transactions on Communications, vol. 64, no. 2, pp. 618-629, Feb. 2016.

[19] J. Östman, G. Durisi, E. G. Ström, M. C. Coşkun and G. Liva, "Short Packets Over Block-Memoryless Fading Channels: Pilot-Assisted or Noncoherent Transmission?," in IEEE Transactions on Communications, vol. 67, no. 2, pp. 1521-1536, Feb. 2019.

[20] A. Martinez and A. G. i Fàbregas, "Saddlepoint approximation of random-coding bounds," 2011 Information Theory and Applications Workshop, La Jolla, CA, USA, 2011, pp. 1-6.

[21] T. Erseghe, "Coding in the Finite-Blocklength Regime: Bounds Based on Laplace Integrals and Their Asymptotic Approximations," in IEEE Transactions on Information Theory, vol. 62, no. 12, pp. 6854-6883, Dec. 2016.

[22] 3GPP TS 38.212 V16.2.0, " Technical Specification Group Radio Access Network, Multiplexing and channel coding", July 2020.

[23] E. Arikan, "Channel Polarization: A Method for constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels", IEEE Transactions on Information Theory, vol. 55, No. 7, pp. 3051-3073, July 2009.

[24] A. Balatsoukas-Stimming, M. B. Parizi and A. Burg, "LLR-Based Successive Cancellation List Decoding of Polar Codes," in IEEE Transactions on Signal Processing, vol. 63, no. 19, pp. 5165-5179, Oct.1, 2015.

[25] I. Tal and A. Vardy, "List Decoding of Polar Codes," in IEEE Transactions on Information Theory, vol. 61, no. 5, pp. 2213-2226, May 2015.

[26] Q. Zhang, A. Liu, X. Pan and K. Pan, "CRC Code Design for List Decoding of Polar Codes," in IEEE Communications Letters, vol. 21, no. 6, pp. 1229-1232, June 2017.

[27] K. Niu and K. Chen, "CRC-Aided Decoding of Polar Codes", IEEE Communications Letters, vol. 16, No. 10, pp.1668-167, October 2015.

[28] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes," 2008 IEEE International Symposium on Information Theory, Toronto, ON, Canada, 2008, pp. 1173-1177.

[29] U. U. Fayyaz and J. R. Barry, "Polar codes for partial response channels," 2013 IEEE International Conference on Communications (ICC), Budapest, Hungary, 2013, pp. 4337-4341.

[30] R. G. Gallager, "Low Density Parity Check Codes", Cambridge, USA: MIT Press, July 1963.

[31] T. Nguyen, T. N. Tan, and H. Lee, "Efficient QC-LDPC Encoder for 5G New Radio", Electronics, vol. 8, p. 668, June 2019.

[32] D. E. Hocevar, "A reduced complexity decoder architecture via layered decoding of LDPC codes," IEEE Workshop onSignal Processing Systems, 2004. SIPS 2004., Austin, TX, USA, 2004, pp. 107-112.

[33] J. Zhang, Y. Lei, and Y. Jin, "Check-node lazy scheduling approach for layered belief propagation decoding algorithm," Electron. Lett., vol. 50, no. 4, pp. 278–279, Feb. 2014.

[34] LIFANG WANG "Implementation of Low-Density Parity-Check codes for 5G NR shared channels," master Thesis, August 30, 2021.

[35] 3GPP TS 38.211 V16.5.0, " Technical Specification Group Radio Access Network, Physical Channels and Modulation", May 2021.

[36] J. R. Hampton, "Introduction to MIMO Communications," Cambridge University Press, Nov. 2013.

[37] R. W. Heath Jr., A. Lozano, "Foundations of MIMO Communication," Cambridge University Press, Dec. 2018.

[38] Robert G. Gallager, "Circularly-Symmetric Gaussian random vectors,"Massachusetts Institute of Technology, January 1, 2008.

[39] Sylvester, James J., "On the relation between the minor determinants of linearly equivalent quadratic functions". Philosophical Magazine 1, pp. 295–305, 1851.

[40] Woodbury, Max A., "Inverting modified matrices", Statistical Research Group, Memo. Rep. no. 42, Princeton University, Princeton, N. J., 1950. pp.4.

[41] G. Fubini, "Sugli integrali multipli", Rom. Acc. L. Rend. (5), vol. 16, no. 1, pp. 608-614, 1907.

[42] J. Jiao, K. Liang, B. Feng, Y. Wang, S. Wu and Q. Zhang, "Joint Channel Estimation and Decoding for Polar Coded SCMA System Over Fading Channels," in IEEE Transactions on Cognitive Communications and Networking, vol. 7, no. 1, pp. 210-221, March 2021.

[43] I. S. Gradshteyn and I. M. Ryzhik, "Table of Integrals, Series and Products" Academic Press, Seventh edition, 2007.