Enhanced Low-Complexity Receiver Design for Short Block Transmission Systems

Mody Sy 🆔 and Raymond Knopp 🆔

EURECOM, 06410 BIOT, France

mody.sy@eurecom.fr, raymond.knopp@eurecom.fr

Abstract—This paper presents a comprehensive analysis and the performance enhancement of short block length channel detection incorporating training information. The current communication systems' short block length channel detection are assumed to typically consist of least squares channel estimation, followed by quasi-coherent detection. By investigating the receiver structure, specifically the estimator-correlator, we show that the non-coherent term, which is often disregarded in conventional detection metrics, results in significant losses in terms of performance and sensitivity in typical operating regimes of 5G/6G systems. A comparison with the fully non-coherent receiver in multi-antenna configurations reveals substantial losses in low spectral efficiency operating areas. Additionally, we demonstrate that by employing an adaptive DMRS/data power adjustment, it is possible to reduce the performance loss gap which is amenable to a more sensitive quasi-coherent receiver. However, both of the aforementioned ML detection strategies can result in substantial computational complexity when processing long bit length codes. We propose an approach to tackle this challenge by introducing the principle of block/segment coding using First-Order RM Codes which is amenable to low-cost decoding through block-based fast Hadamard transforms. The Block-based FHT has demonstrated to be cost-efficient with regards to decoding time, as it evolves from quadric to quasi-linear complexity with a manageable decline in performance. Additionally, by incorporating an adaptive DMRS/data power adjustment technique, we are able to bridge/reduce the performance gap with respect to the conventional maximum likelihood receiver and attain high sensitivity, leading to a good trade-off between performance and complexity to efficiently handle small payloads.

Index Terms—5G NR, Short data Transmission, Reed Muller codes, Maximum Likelihood Decoding, Decoding via Fast Hadamard Transform.

I. INTRODUCTION

The New Radio (NR) waveforms, such as *Physical Uplink Control Channel (PUCCH)*, have been specifically designed to transmit small payloads with minimal error rates in challenging signal-to-noise conditions. To fully leverage their potential for *Ultra-Reliable Low Latency Communications* (URLLC) and their suitability for massive machine-type communications, it is crucial to develop robust coding strategies that enable low-complexity detection/decoding algorithms.

Therefore, there is a need for enhanced receiver designs that can accurately and reliably detect short data transmissions with low computational complexity and power consumption [1]. The area of short block transmission has been thoroughly examined in the existing literature, covering various aspects such as the design of signal codes and the derivation of state-of-the-art converse and achievability bounds for both coherent and non-coherent communication [2]–[5].

This study aims to explore the enhancement of short-block length detection/decoding strategies. The main focus of this work is the baseline 3GPP PUCCH transmission, specifically the use of Reed-Muller codes paired with *Orthogonal De-*

modulation Reference Signals (DMRS). The symbols, being message-independent, are used by the receiver to resolve channel uncertainty through explicit channel estimation or more advanced joint estimation and detection techniques. Furthermore, through the structure of the receiver, namely the *correlator-estimator*, we will show that the non-coherent term which is not typically used in conventional receivers can lead to a performance penalty which can have a significant effect on receiver sensitivity in the operating regimes of 5G/6G system. Additionally, in current communications systems, short block lengths detection/decoding usually consists of least-squares channel estimation followed by quasi- coherent detection, but it should be noted that in both cases, the underlying algorithms are sub-optimal because of the detection procedure involving separate channel estimation and *quasi-coherent detection*. Moreover, a *fully non-coherent receiver* also incurs a substantial complexity cost, particularly when dealing with longer bit length transmissions, despite providing enhanced or acceptable detection performance in comparison to a *conventional receiver*. Therefore, it becomes realistic to consider alternative detection/decoding strategies that offer a favorable performance/complexity trade-off.

It is noteworthy that the use of *maximum likelihood* (ML) decoding for the above-mentioned receivers seems computationally demanding. Therefore, we propose an approach to tackle this challenge by introducing the principle of block/segment encoding using First-Order RM Codes which is amenable to low-cost decoding through block-based *Fast Hadamard Transforms* (FHT). The Block-based FHT has demonstrated to be cost-efficient with regards to decoding time, as it evolves from quadric to quasi-linear complexity with a manageable decline in performance. Additionally, by incorporating an adaptive DMRS/data power adjustment technique, it is possible to bridge/reduce the performance gap with respect to the conventional maximum likelihood receiver and attain high sensitivity, leading to a good trade-off between performance and complexity to efficiently handle small payloads.

The manuscript is structured as follows: Section II lays out the system model, the methodology is drawn up in Section III, Section IV presents the results and performance analysis, and finally Section V concludes the paper.

II. SYSTEM MODEL

Consider a discrete-time model in which the transmitted and received symbols are N-dimensional column vectors, and thus a system is designed in such a way that the relationship between the transmitted and received signals at the $i - th$ antenna port is as follows:

$$\mathbf{y}_i = \mathbf{h}_i \mathbf{x} + \mathbf{z}_i, \quad i = 0, 1, \ldots, \mathsf{N_R} - 1, \qquad (1)$$

where $\mathbf{y}_i$ represents an observed vector in N complex dimensions, $\mathbf{x}$ is an N-dimensional modulated vector transporting K channel bits, so that the message $\mathcal{M} = 0, 1 \ldots, 2^K - 1$, $\mathbf{z}_r$ is additive white Gaussian noise whose real and imaginary components are independent and have variance σ^2 in each dimension. $\mathsf{N_R}$ represents the number of observations of the transmitted vector over a multi-antenna receiver. The transmitted vector $\mathbf{x}$ typically consists of data-independent components known as pilot or reference signals, so that $\mathbf{x} = \mathbf{x}^{(\mathsf{d})} + \mathbf{x}^{(\mathsf{p})}$. The subscripts (d) and (p) serve to denote the data components and reference signals respectively. These reference signals serve to mitigate channel ambiguity in time, frequency, and space and are used to estimate the channel. In practice, the reference signals are *interleaved* among the data-dependent components to account for the characteristics of the propagation channel. This interleaving technique is commonly used in current OFDM systems. 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, the standard notation for N is $12\mathcal{PL}$, where $\mathcal{P}$ refers to the number of *physical resource blocks* or (PRBs), each consisting of 12 complex dimensions or resource elements. The value of $\mathcal{P}$ usually falls within a range of 1 to 16. $\mathcal{L}$ represents the number of symbols, typically ranging from 1 to 14, but it can be increased if multiple slots are utilized to signal the K bits. To demonstrate the consequences of disregarding the non-coherent term, PUCCH format 2 will be employed as an illustration. This particular format, based on OFDM, is a short PUCCH format and has the capability to transmit more than two bits utilizing either one or two OFDM symbols [6]. The transmission framework exhibits a simple procedure. Prior to transmitting short block channels, ranging in length from 3 to 11 bits, the messages are subjected to encoding utilizing a $(32, K)$ Reed-Muller coding scheme. $c_\ell = \left(\sum_{k=0}^{K-1} b_k \cdot M_{\ell,k} \right) \bmod 2$, where $\ell = 0, 1, \ldots, N' - 1$, $N' = 32$ and $M_{\ell,k}$ represents the basis sequences as defined in [7]. It turns out that a classical (N', K) block code is utilized among the family of Reed-Muller codes with a block length of 32, denoted as N'. The information bitwidth is defined as K. The input bit sequence to the encoder is $b(0), b(1), b(2), \ldots, b(K-1)$ and the resulting sequence before rate matching is represented as $c(0), c(1), c(2), \ldots, c(N' - 1)$. The output bit sequence, following rate matching, is denoted as $e(0), e(1), e(2), \ldots, e(E-1)$, where the length of the rate matching output sequence, E, is dependent on the number of PRBs. The rate matching process in this case is simply repetition. The encoded *Uplink Control Information* (UCI) payload, denoted as $\mathbf{e}$, undergoes a scrambling process prior to modulation, resulting in a new block of bits $\tilde{\mathbf{e}}$ such that $\tilde{e}(\ell) = [e(\ell) + d(\ell)] \bmod 2$. The scrambling sequence $d(\ell)$ is a pseudo-random sequence derived from *radio resource control* (RRC) configuration parameters. The $\tilde{\mathbf{e}}$ bits are then subjected to *quadrature phase shift keying* (QPSK) modulation, resulting in a block of complex-valued modulation symbols $x(0), x(1), \ldots, x(E/2 - 1)$. The resource mapping process follows, whose aim is to allocate the modulated symbols onto resource occasions in both time and frequency domains and inserting DMRS resources. As illustrated in Figure 1, the resource mapping here is embedded in the same spirit as in 3GPP PUCCH2 transmission. On reception, the messages are
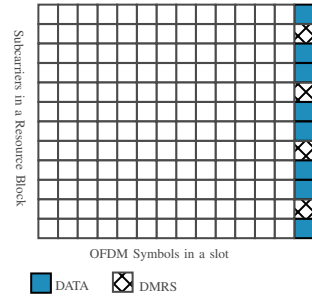

Fig. 1: General resource mapping.

recovered using maximum likelihood decoding, which can be computationally demanding.

III. METHODOLOGY

A. Maximum Likelihood Receivers

We describe detection metric for classical non-coherent channel with unknown phase. The LOS channel is $\mathbf{h}_i = \exp(j\theta_i)\mathbf{I}$ where θ_i are assumed to be i.i.d. uniform random variables on $[0, 2\pi)$. We denote the likelihood function for the observed vector on a particular receiver branch with respect to a given transmitted signal as

$$q(\mathbf{x}, \{\mathbf{y}_i\}) = p(\{\mathbf{y}_i\} \mid \mathbf{x}) = p(\{\mathbf{y}_i\} \mid \mathbf{x}, \theta_i) p(\theta_i), \quad (2)$$

where we have the fact que $p(\theta_i) = \frac{1}{2\pi}$, is a uniform distribution on $[0, 2\pi)$. The likelihood function is equivalent to

$$q(\mathbf{x}, \{\mathbf{y}\}) = \prod_{i=0}^{\mathsf{N_R}-1} \frac{1}{2\pi} \int_{\theta_i=0}^{2\pi} \exp\left(-\frac{||\mathbf{y}_i - \mathbf{h}_i\mathbf{x}||^2}{N_0} \right) \mathrm{d}\theta_i. \quad (3)$$

Moreover we have

$$||\mathbf{y}_i - \mathbf{h}_i\mathbf{x}||^2 = ||\mathbf{y}_i||^2 + ||\mathbf{h}_i\mathbf{x}||^2 - 2\mathrm{Re}(\mathbf{y}_i\mathbf{h}_i^*\mathbf{x}^*). \quad (4)$$

Considering that $\mathbf{x}^\dagger\mathbf{y}_i = |\mathbf{x}^\dagger\mathbf{y}_i| e^{j\phi_i}$ where $\phi_i = \angle\mathbf{x}^\dagger\mathbf{y}_i$, the likelihood function subsequent to the exclusion of multiplicative terms that are unrelated to $\mathbf{x}$ can be expressed as:

$$q(\mathbf{x}, \{\mathbf{y}\}) \propto \prod_{i=0}^{\mathsf{N_R}-1} \exp\left(-\frac{||\mathbf{x}||^2}{N_0} \right) \cdot$$
$$\frac{1}{2\pi} \int_0^{2\pi} \exp\left(\frac{2}{N_0} |\mathbf{x}^\dagger\mathbf{y}_i| \cos(\phi_i - \theta_i) \right) \mathrm{d}\theta_i, \quad (5)$$

where $\frac{1}{\pi} \int_{\lambda=0}^{\pi} \exp(z\cos(\lambda))\mathrm{d}\lambda = \mathrm{I}_0(z)$, $\mathrm{I}_0(\cdot)$ is the modified bessel function of the first kind and $\dagger$ denote the complex conjugate transpose or Hermitian.

$$q(\mathbf{x}, \{\mathbf{y}\}) = \prod_{i=0}^{\mathsf{N_R}-1} \exp\left(-\frac{||\mathbf{x}||^2}{N_0} \right) \mathrm{I}_0\left(\frac{2}{N_0} |\mathbf{x}^\dagger\mathbf{y}_i| \right). \quad (6)$$

We typically simplify the Likelihood fonction from (6) via a *max-log approximation*. The approximation is obtained by assuming that $\log\{\sum_i \exp(\lambda_i)\} \sim \max i \{\lambda i\}$, taking into account the exponential approximation of the modified Bessel function of the first kind, $\mathrm{I}_0(z) \sim \frac{e^z}{\sqrt{2\pi z}} \sim e^z$.

Hence, the maximum likelihood detection metric of the input signal can be closely approximated as:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\arg\max} \left(\sum_{i=0}^{N_R-1} \frac{2}{N_0} \left| \mathbf{x}^\dagger \mathbf{y}_i \right| - \frac{||\mathbf{x}||^2}{N_0} \right). \qquad (7)$$

It is worth mentioning that, in (7), many of the terms can be omitted when $||\mathbf{x}||$ is constant, as would be the case for BPSK or QPSK modulation for instance.

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\arg\max} \sum_{i=0}^{N_R-1} \left| \mathbf{x}^\dagger \mathbf{y}_i \right| = \underset{\mathbf{x}}{\arg\max} \sum_{i=0}^{N_R-1} \left| \mathbf{x}^\dagger \mathbf{y}_i \right|^2. \qquad (8)$$

The form of the metrics has been observed to be distinct, with no overlap between the data and DMRS symbols. By writing $\mathbf{x} = \mathbf{x}^{(p)} + \mathbf{x}^{(d)}$, then

$$\left| \left(\mathbf{x}^{(d)\dagger} + \mathbf{x}^{(p)\dagger} \right) \mathbf{y}_i \right|^2 = \left| \mathbf{x}^{(p)\dagger} \mathbf{y}_i^{(p)} + \mathbf{x}^{(d)\dagger} \mathbf{y}_i^{(d)} \right|^2 \qquad (9)$$

More generally, in the traditional literature, the detection of signals with training information can be extended using (8) which is often referred to as an estimator-correlation. The equation encompasses both a quasi-coherent detection term and a non-coherent energy term.

$$\hat{\mathbf{x}}^{(d)} = \underset{\mathbf{x}^{(d)}}{\arg\max} \underbrace{\sum_{i=0}^{N_R-1} \left| \mathbf{x}^{(p)\dagger} \mathbf{y}_i^{(p)} \right|^2}_{\text{data-independent term}} + \underbrace{\sum_{i=0}^{N_R-1} \left| \mathbf{x}^{(d)\dagger} \mathbf{y}_i^{(d)} \right|^2}_{\text{non-coherent data term}} +$$

$$\underbrace{2 \operatorname{Re} \left(\sum_{i=0}^{N_R-1} \mathbf{x}^{(p)\dagger} \mathbf{y}_i^{(p)} \cdot \mathbf{x}^{(d)} \mathbf{y}_i^{(d)\dagger} \right)}_{\text{quasi-coherent term}},$$

$$(10)$$

where $\mathbf{x}^{(p)\dagger} \mathbf{y}_i^{(p)} = \left(\mathbf{x}^{(p)\dagger} \mathbf{x}^{(p)} \right) \hat{\mathbf{h}}_i = N_p \hat{\mathbf{h}}_i$.

We demonstrate that neglecting the non-coherent energy term in current systems leads to a significant performance penalty in typical operating regimes of 5G/6G systems. This is contrary to the widespread belief that the impact of non-coherent energy term is negligible.

B. Low Complexity Receiver Design

Reed-Muller codes (RM codes) are commonly known to be decodable using Hadamard or Fast Hadamard transforms. However, it is well-established that decoding the first-order RM code ($RM(r = 1, m)$) using FHT is easier compared to higher-order RM codes ($r \geq 2$). In recent literature, several innovative algorithms have been proposed for decoding RM codes of any order [8]–[12].

Although maximum likelihood decoding algorithms have been extensively investigated in traditional literature for decoding data packets encoded with first-order Reed-Muller codes [13], it can become computationally expensive when the message length exceeds 6 bits. This is because the resulting codewords tend to be excessively long, leading to complex decoding processes that involve high-dimensional Hadamard transforms when using a FHT-based decoder. This presents a significant challenge for transmitting short packets as the cost can be substantial. As an illustration, for a message of $K = 11$ bits, the length of the code words would be $N' = 2^{10}$ bits using a first order $RM(1, m = 10)$. Hence, to address the constraint of having a message length of $K \geq 6$ bits, we can utilize the principle of encoding and decoding by blocks. This method

takes advantage of the low complexity decoding offered by FHT-based decoders. The objective is to segment the message into smaller, more manageable segments of bits, which can then be fed into $RM(1, m)$ encoders and concatenated. Upon reception, the received code is deconcatenated and decoded through the appropriate dimension of the Hadamard transform which is amenable to a low complexity receiver.

1) Block-based Encoding Principle:

In instances where the payload exceeds 6 bits, such as in the case of $K = 11$ bits, a combination of two first-order RM codes, $RM(1, m = 4)$ and $RM(1, m = 5)$, can be employed to encode the respective sub-blocks of 5 bits and 6 bits.

In regards to the Reed-Muller code $RM(1, 4)$, the codewords are generated using (11)

$$\mathbf{c}^{(1)} = \mathbf{m}^{(1)} \mathbf{G}^{(1)} = \mathbf{m}^{(1)} \cdot \begin{bmatrix} \mathbf{1} & \mathbf{v}_4 & \mathbf{v}_3 & \mathbf{v}_2 & \mathbf{v}_1 \end{bmatrix}^{\mathrm{T}}.$$
$$(11)$$

This code is characterized as an $(N' = 16, K = 5, d_{min} = 8)$ code, where the minimum distance of the $RM(r, m)$ is defined as 2^{m-r}. The monomials of degree less than or equal to r are represented by $\mathbf{1}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$, with associated vectors, as indicated in the table below.

1	$\mathbf{v}_4$	$\mathbf{v}_3$	$\mathbf{v}_2$	$\mathbf{v}_1$
1	0	0	0	0
1	0	0	0	1
1	0	0	1	0
1	0	0	1	1
1	0	1	0	0
1	0	1	0	1
1	0	1	1	0
1	0	1	1	1
1	1	0	0	0
1	1	0	0	1
1	1	0	1	0
1	1	0	1	1
1	1	1	0	0
1	1	1	0	1
1	1	1	1	0
1	1	1	1	1

If we consider $RM(1, 5)$, code words are generated by

$$\mathbf{c}^{(2)} = \mathbf{m}^{(2)} \mathbf{G}^{(2)} = \mathbf{m}^{(2)} \cdot \begin{bmatrix} \mathbf{1} & \mathbf{v}_5 & \mathbf{v}_4 & \mathbf{v}_3 & \mathbf{v}_2 & \mathbf{v}_1 \end{bmatrix}^{\mathrm{T}}.$$
$$(12)$$

This is a $(N' = 32, K = 6, d_{min} = 16)$ code. Therefore, the process of concatenation entails the merging of the two sub-codes,

$$\mathbf{c} = [\mathbf{c}^{(1)} \ \mathbf{c}^{(2)}] = [\mathbf{m}^{(1)} \mathbf{G}^{(1)} \ \mathbf{m}^{(2)} \mathbf{G}^{(2)}]. \qquad (13)$$

2) Block-based Decoding via FHT:

Consider the received sequence $\mathbf{u} = (u_0, u_1, \ldots, u_{2^m-1}) \in \mathbb{F}_2$, and let $\mathbf{c} = (c_0, c_1, \ldots, c_{2^m-1}) \in \mathbb{F}_2$ be a codeword. The bipolar representation of $\mathbf{u}$ is denoted as $\mathbf{U} \in \{-1, +1\}$ and is defined as $\mathbf{U} = (-1)^{\mathbf{u}}$. $\mathbb{F}_2$ denotes the Galois field. Similarly, the bipolar representation of $\mathbf{c}$ is denoted as $\mathbf{C}$ and defined as $\mathbf{C} = (-1)^{\mathbf{c}}$. The decoding algorithm involves computing the correlation between $\mathbf{U}$ and $\mathbf{C}_i$, denoted as Δ_i, for each of the 2^{m-1} codewords $\mathbf{C}_i = (-1)^{\mathbf{c}_i}$. The final step is to select the codeword for which Δ_i is the maximum. The simultaneous computation of all correlations can be depicted as a matrix representation. Denoting the column vector C_i and constructing the matrix $\mathbf{H} = \begin{bmatrix} \mathbf{C}_0 & \mathbf{C}_1 & \ldots & \mathbf{C}_{2^m-1} \end{bmatrix}$, the computation of all correlations can be expressed as follows:

$$\boldsymbol{\Delta} = \mathbf{U}\mathbf{H}. \qquad (14)$$

Where $\mathbf{H}$ is a Hadamard matrix of dimension 2^m. For the first subblock, which utilizes a $RM(1,4)$ code with generator matrix $\mathbf{G}^{(1)}$, $\mathbf{H}_{16}^{(1)}$ is employed. Similarly, for the second subblock, which employs a $RM(1,5)$ code with generator matrix $\mathbf{G}^{(2)}$, $\mathbf{H}_{32}^{(2)}$ is utilized.

Furthermore, comprehensive descriptions of algorithms for first order RM code decoding using the Hadamard transform can be found in the works of Moon [14] and Wicker [15]. This decoding process can be optimized through the utilization of a FHT which is applicable to Hadamard matrices produced through the Sylvester construction. This optimization is based on the fact that $\mathbf{H}_{2^m} = \mathbf{H}_2 \otimes \mathbf{H}_{2^{m-1}}$, where the Kronecker product of matrices, denoted by $\otimes$, is applied. As a result, the matrix $\mathbf{H}_{2^m}$ can be decomposed as stated in the theorem derived from linear algebraic principles [14].

$$\mathbf{H}_{2^m} = \mathbf{W}_{2^m}^{(1)} \mathbf{W}_{2^m}^{(2)} \cdots \mathbf{W}_{2^m}^{(m)}. \tag{15}$$

where $\mathbf{W}_{2^m}^{(i)} = \mathbf{I}_{2^{m-i}} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^{i-1}}$, $\mathbf{I}$ is an identity matrix. Thus it comes,

$$\begin{aligned} \mathbf{H}_{16}^{(1)} &= \mathbf{W}_{16}^{(1)} \mathbf{W}_{16}^{(2)} \mathbf{W}_{16}^{(3)} \mathbf{W}_{16}^{(4)} \\ &= \left(\mathbf{I}_{2^3} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^0}\right)\left(\mathbf{I}_{2^2} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^1}\right) \\ &\quad \left(\mathbf{I}_{2^1} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^2}\right)\left(\mathbf{I}_{2^0} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^3}\right). \end{aligned} \tag{16}$$

Let's consider $\mathbf{U}^{(1)} = [U_0, U_1, \ldots, U_{15}]$, the first received sequence to be fed to the first decoder. The corresponding Hadamard transforms can then be written as

$$\boldsymbol{\Delta}^{(1)} = \mathbf{U}^{(1)} H_{16}^{(1)} = \mathbf{U}^{(1)}\left(\mathbf{W}_{16}^{(1)} \mathbf{W}_{16}^{(2)} \mathbf{W}_{16}^{(3)} \mathbf{W}_{16}^{(4)}\right), \tag{17}$$

where

$$\begin{aligned} \mathbf{W}_{16}^{(1)} &= \mathbf{I}_8 \otimes \mathbf{H}_2 \\ \mathbf{W}_{16}^{(2)} &= \mathbf{I}_4 \otimes \mathbf{H}_2 \otimes \mathbf{I}_2 \\ \mathbf{W}_{16}^{(3)} &= \mathbf{I}_2 \otimes \mathbf{H}_2 \otimes \mathbf{I}_4 \\ \mathbf{W}_{16}^{(4)} &= \mathbf{H}_2 \otimes \mathbf{I}_8. \end{aligned}$$

The conventional computation of the Hadamard transform $\mathbf{H}_{2^m}$ results in 2^m elements, each of which is obtained through 2^m addition/subtraction operations. This leads to a computational complexity of $(2^m)^2 = \mathcal{O}\left(N'^2\right)$, which is equivalent to the complexity of a standard ML decoder that operates in a quadratic order. In contrast, the FHT, which has m stages, has a computational complexity of $m2^m = \mathcal{O}\left(N' \log N'\right)$ (i.e., *quasi-linear complexity*) due to its 2^m addition/subtraction operations per stage. The procedure of block-based RM decoding is therefore depicted in Figure 2.

IV. NUMERICAL RESULTS

For illustration purposes, we focus on PUCCH2-based short block lengths. PUCCH2 is configurable in terms of the resource usage, but we consider the simplest comprising of 2 groups of 12 dimensions or *resource elements*, so-called PRBs, making 24 dimensions which consist of 16 for data components, and 8 for DMRSs, which are known symbols used for channel estimation and tracking. The simulations were performed under the assumption of TDL-C Non-Line-Of-Sight 300 ns wireless channel, utilizing 2, 4 and 8 antenna configurations. The TDL-C channel model, which adheres to the 3GPP reference specifications, was employed. This model is distinguished by its extended time delay spread and its

primary focus on non-MIMO assessments [16]. The antenna ports were subjected to independent and identically distributed realizations, with no incorporation of correlation modeling.

In Figure 3a, the comparison of performance between the optimal ML receiver, which is a *fully non-coherent detector*, and the conventional receiver, which is a quasi-coherent detector, is presented. The difference in performance arises due to the fact that the *non-coherent energy term* is disregarded in the quasi-coherent detector.

In Figure 3a, it is demonstrated that as the BLER threshold is set to 1%, the performance losses incurred by the penalties increase. This is evidenced by the estimated performance losses of 0.2 dB, 0.6 dB and 1.8 dB for the 2, 4 and 8 receive antenna configurations, respectively. Furthermore, it can be deduced that the performance gap expands with the increase in the number of receiving antennas. This reflects that operating in regions of low spectral efficiency while disregarding the *non-coherent term* leads to a significant degradation in performance and a decrease in receiver sensitivity. These regions are defined by a low signal-to-noise ratio (SNR), which greatly affects the precision of channel estimates in each receiving chain. Consequently, the significance of the *non-coherent energy term* becomes increasingly evident. From this insight, the quasi-coherent receiver is sub-optimal in terms of its performance, despite its reduced complexity.
The observed decline in performance could potentially be remediated through the enhancement of channel estimate accuracy.
Furthermore, it is also conceivable to address the performance discrepancies observed in comparison to the *fully non-coherent receiver* by employing an adaptive DMRS/data power adjustment. In instances where we transmit reference and data symbols jointly in common OFDM symbols, as is the case for PUCCH or PUSCH or in some downlink control channels, there lies the potential to adjust the power ratio between data and reference signals. This is applicable to both downlink and uplink transmissions without incurring any penalty in terms of peak power increase. Therefore, the system can be conceived as $\mathbf{y}_i = \left(\mathbf{x}^{(d)} + \beta \cdot \mathbf{x}^{(p)}\right)\mathbf{h}_i + \mathbf{z}_i$. This adaptive power adjustment procedure is contingent on the values of β. The DMRS Power is to be slightly increased in a judicious manner since β must be perfectly calibrated to ensure compliance with potential radio frequency constraints. As depicted in Figure 3b, the performance improvement can be observed as a function of varying values of β. Optimal performance improvement is achieved when β is set to 1.75. As demonstrated through simulation, when nRx=4 and β is selected as 1.5, the sub-optimal ML receiver exhibits comparable performance to the optimal ML receiver, effectively bridging the initial 0.6 dB gap at the *Bit Error Rate* (BLER) threshold of 1%. Furthermore, as we stated earlier, message recovery using maximum likelihood decoding is computationally demanding for both *fully non-coherent* and *quasi-coherent* receivers. Therefore, through Figure 4, we provide some simulation results to show the merit of the proposed block-based decoding via HT/FHT with respect to the recovery of information bits using a quasi-coherent ML decoder. The results suggest that the ML decoder exhibits improved performance, however, this advantage is accompanied by a significant increase in computational complexity, even when utilizing the block decoding principle. In practical implementation, each received
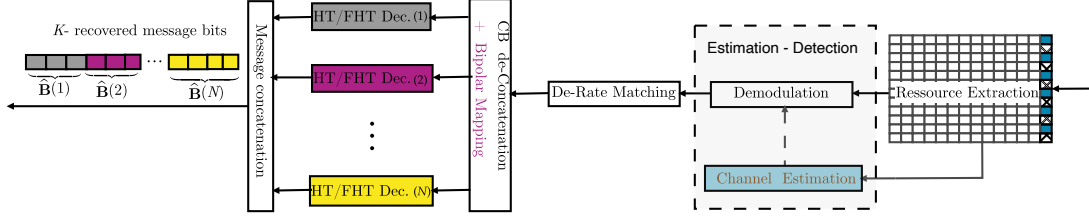
Fig. 2: Block-based HT/FHT based-decoding of Short blocks : Receiver end.
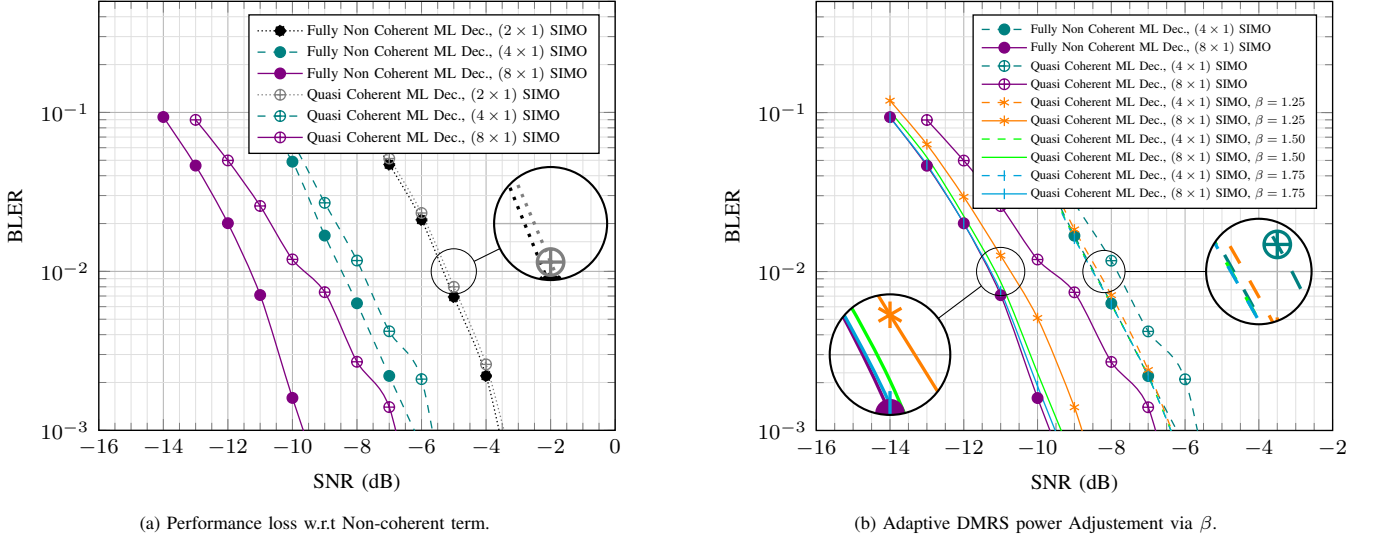

(a) Performance loss w.r.t Non-coherent term.

(b) Adaptive DMRS power Adjustement via β.

Fig. 3: Block Error Rate, 4 bits, ($\{8, 4, 2\} \times 1$) SIMO, 3GPP TDL-C urban NLOS channel, 300ns of Delay spread, Unknown *Channel State Information*.
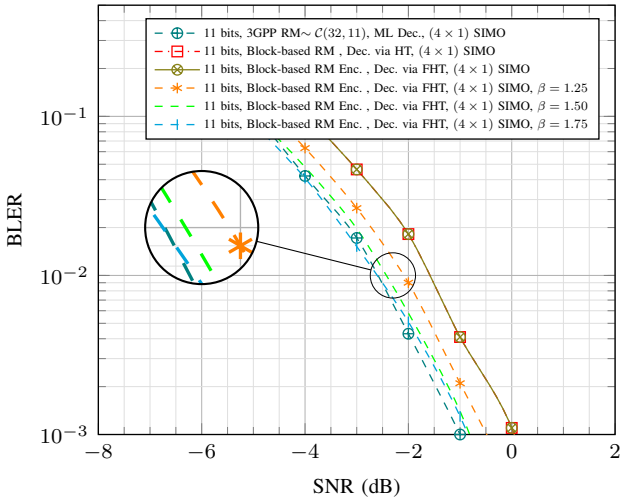

Fig. 4: Block Error Rate, 11 bits, (4×1) SIMO, 3GPP TDL-C urban NLOS channel, 300ns of Delay spread, Unknown *Channel State Information*, Adaptive DMRS power Adjustement via β.

code block ($\mathcal{C}(11, 48) = [\mathcal{C}_1(5, 16), \mathcal{C}_2(6, 32))]$) is processed through a corresponding decoder, followed by concatenation at the decoder output stage. It is important to acknowledge that the efficacy of block decoding through both HT and FHT is comparable, exhibiting a marginal deviation of 1 dB from the *Maximum Likelihood* (ML) receiver at a BLER threshold of 0.1%. In terms of computational efficiency, block decoding via Hadamard transform demonstrates superior performance with a more expeditious convergence time. Consequently, it is advisable to employ this technique in scenarios where time sensitivity is a critical factor, specifically in URLLCs and

sporadic random access applications.

Additionally, the performance disparity between the ML receiver and FHT based-receiver can be mitigated via the adaptive power adjustment procedure decribed above. When the number of receive antennas is 4 and β is selected as $\beta = 1.75$, the performance of the FHT-based decoder approaches that of the ML decoder at a BLER threshold of 1%, which corresponds to 1 dB in terms of additional gain. This result is significant as it indicates that the FHT-based decoder provides a better trade-off between complexity and performance for short packet decoding, owing to its lower computational complexity. Overall, the implications of perfoming an adaptive DMRS/power ajustement within the 3GPP standard are significant in term of performance improvement. Specifically, the *User Equipment* (UE) could vary the power allocation between the DMRS and data. This flexibility in power allocation is transparent to the receiver.

V. CONCLUSIONS

In this paper, we presented a comprehensive analysis and the performance enhancement of short block length channel detection incorporating training information. By investigating the receiver structure, specifically the estimator-correlator, it was shown that the non-coherent term, which is often disregarded in conventional detection metrics, results in significant losses in terms of performance and sensitivity in typical operating regimes of 5G/6G systems. Additionally, it was demonstrated that by employing an adaptive DMRS/data power adjustment, it is possible to reduce the performance loss gap which is amenable to a more sensitive and low complexity quasi-coherent receiver. However, both of the aforementioned

ML detection strategies could result in substantial computational complexity when processing long bit length codes. We proposed an innovative approach to tackle this challenge by introducing the principle of block/segment coding using First-Order RM Codes which is amenable to low-cost decoding through block-based fast Hadamard transforms. Thus, by incorporating an adaptive DMRS/data power adjustment technique, we were able to bridge/reduce the performance gap with respect to the conventional maximum likelihood receiver and attain high sensitivity, leading to a good trade-off between performance and complexity to efficiently handle small payloads.

REFERENCES

[1] 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.

[2] 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. 810-81.

[3] 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.

[4] 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.

[5] 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.

[6] 3GPP TS 38.211 V16.5.0 ," Technical Specification Group Radio Access Network, Physical Channels and Modulation", May 2021.

[7] 3GPP TS 38.212 V16.5.0 ," Technical Specification Group Radio Access Network, Multiplexing and channel coding", July 2020.

[8] I. Dumer and K. Shabunov, "Soft-decision decoding of Reed-Muller codes: recursive lists," in IEEE Transactions on Information Theory, vol. 52, no. 3, pp. 1260-1266, March 2006.

[9] M. Ye and E. Abbe, "Recursive projection-aggregation decoding of Reed-Muller codes," 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France, 2019, pp. 2064-2068.

[10] J. Li, S. M. Abbas, T. Tonnellier and W. J. Gross, "Reduced Complexity RPA Decoder for Reed-Muller Codes," 2021 11th International Symposium on Topics in Coding (ISTC), Montreal, QC, Canada, 2021, pp. 1-5.

[11] M. Lian, C. Häger and H. D. Pfister, "Decoding Reed–Muller Codes Using Redundant Code Constraints," 2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 2020, pp. 42-47.

[12] N. Doan, S. A. Hashemi, M. Mondelli and W. J. Gross, "Decoding Reed-Muller Codes With Successive Codeword Permutations," in IEEE Transactions on Communications, vol. 70, no. 11, pp. 7134-7145, Nov. 2022.

[13] Alexei E. Ashikhmin and Simon N. Litsyn, "Fast Decoding Algorithms for First Order Reed-Muller and Related Codes". Designs, Codes and Cryptography 7, pp. 187–214, 1996.

[14] Todd K. Moon,"Error Correction Coding: Mathematical Methods and Algorithms", Wiley-Interscience605 Third Avenue New York, NYUnited States, May 2005

[15] Stephen B. Wicker,"Error Control Systems for Digital Communication and Storage", Prentice Hall, July 1994.

[16] 3GPP TR 38 901 v16.1.0," Technical Report, Study on channel model for frequencies from 0.5 to 100 GHz", November 2020.