DOCTORAL SCHOOL EDITE

École Doctorale Informatique,
Télécommunications et Électronique

Dissertation

Submitted to Sorbonne University
in partial fulfillment of the requirements for the degree of

Ph.D. in Engineering Sciences

Computer Science, Telecommunications, and Electronics

Defended by

Mody Sy

Enhanced Receiver Design for Ultra-Reliable Low-Latency Short Data Communications

---

prepared at EURECOM, Department of Communication systems

defended on October 25th, 2024

Thesis Committee :

Reviewers :	Albert Guillén i Fàbregas	-	Cambridge University	-	United Kingdom
	Giuseppe Durisi	-	Chalmers University of Technology	-	Sweden
Examiners :	Derya Malak	-	EURECOM	-	France
	Catherine Douillard	-	IMT Atlantique	-	France
Advisor :	Raymond Knopp	-	EURECOM	-	France

Acknowledgments

As I bring this PhD journey to a close, I would first like to sincerely thank the entire administrative, academic, and technical staff of EURECOM for allowing me to conduct my research within this remarkable institution.
My deepest thanks go to my supervisor, Professor Raymond Knopp, for his warm welcome, steady guidance, patience, and generous sharing of his knowledge. His support has been a constant source of inspiration.
I am also profoundly grateful to the members of my thesis committee, Professors Albert Guillén i Fàbregas and Giuseppe Durisi, and especially Professors Catherine Douillard and Derya Malak, for their insightful feedback and valuable advice, which were essential to the success of this work.
Furthermore, my most profound thanks are directed to all the professors and educators who have enriched my academic training with their expertise and dedication.
Finally, I take this opportunity to convey my appreciation and profound gratitude to all those who extended a helping hand and brightened my journey. Whether it was through a simple smile or a kind gesture, I thank you from the bottom of my heart.

Thank you.

Dedications

I humbly dedicate this Ph.D. thesis, a fervent testament, to those who have shaped the essence of who I am.

To my beloved parents,

To my wife,

To my brothers and sisters,

To all my family and dear friends.

May you find in this doctoral thesis the vibrant expression of my affection and profound gratitude.

Abstract

Ultra-Reliable Low Latency Communications (URLLC) in the context of 5G and beyond have garnered particular attention in recent years, as underscored by various scholarly works. Numerous efforts are underway to enhance transmission schemes at the physical layer, notably in the design of signal codes and receiver algorithms, to meet reliability and latency requirements. However, although 5G transmission formats can enable very short packet transmission through the use of mini-slots, the ratio of training information to data is not necessarily suitable for extremely short packet transmission. Additionally, short packet 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 the inherent sporadic nature of these transmissions. Simply stated, designing sophisticated and innovative transmission strategies in the wake of short packets is of paramount importance, contributing to shaping the *fourth industrial revolution* (Industry 4.0). Thus, this thesis focuses on exploring important aspects of short data transmission, particularly the design of reliable and low-complexity receivers.

Specifically, we are initially interested in designing advanced BICM receivers for short packets compliant with polar and *low-density parity-check* (LDPC) coded transmissions. Indeed, the relevance of BICM becomes particularly evident in scenarios characterized by error-prone communication channels, requiring increased reliability. Its effectiveness also depends on underlying detection and decoding metrics, highlighting the importance of a nuanced trade-off between performance and complexity. The proposed receiver metrics consider scenarios where *channel state information* (CSI) is effectively unknown, aiming to assess the impact of various channel conditions. Hence, we present enhanced receivers 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 LDPC coded formats. We look into receiver metrics exploiting *joint estimation and detection* (JED), which is amenable to situations where low-density of *demodulation reference signals* (DMRSs) are interleaved with coded data symbols and both are transmitted over a small number of OFDM symbols so that near-perfect channel estimation is impossible. We show that it is possible to enhance the performance and sensitivity through block-wise joint estimation–detection compared to conventional 5G *orthogonal frequency division multiplexing* BICM receivers. Unlike conventional symbol-by-symbol detection in BICM systems, where the observation interval for a particular coded bit is confined to the symbol in which it is conveyed, the proposed method performs block-wise joint detection over a sliding window of adjacent symbols to fundamentally leverages their statistical dependencies. Accordingly, the LLR for a particular coded bit incorporates information from all symbols within the detection window, rather than being constrained to its host symbol alone. Performance evaluation spans SIMO and SU-MIMO configurations, emphasizing the efficacy of the estimation-detection strategy 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 receivers remarkably outperform the conventional ones and can be used to achieve detection performance that is close to that of coherent receivers with perfect CSI. Meanwhile, we look into the impact of DMRS density and how DMRS power boosting may be used to achieve some extra performance.

In addition to this initial contribution, we also explore the aspect of designing reliable and low-complexity receivers compliant with Reed-Muller (RM) coded transmissions, which have regained significant research interest due to their similarity to polar codes and their excellent error correction performance using *maximum-likelihood* (ML) decoding. Moreover, NR waveforms, such as the *physical uplink control channel* (PUCCH), have been specifically designed to transmit small payloads with minimal error rates in challenging signal-to-noise conditions. The design of

enhanced receivers capable of reliably detecting and decoding short packets in the range of 3 to 11 bits with lower complexity and low energy consumption seems more than necessary to fully leverage their potential in URLLC and their use in NR signaling scenarios. Accordingly, we are focusing on the baseline 3GPP PUCCH in DMRS assisted transmission, specifically the use of Reed-Muller codes decodable via ML. Through the receiver structure, namely the *estimator-correlator*, we show that the non-coherent energy term, which is typically not used in conventional receivers, can result in a penalty that can significantly affect receiver performance in the operating regimes of 5G/6G systems. Furthermore, in current communication systems, the detection/decoding of short packets generally consists of channel estimation by the least squares method followed by quasi-coherent detection. However, it should be noted that in both cases, the underlying algorithms are suboptimal due to the detection procedure involving separate channel estimation followed by quasi-coherent detection. Conversely, the use of a *fully non-coherent receiver* would consequently incur a substantial complexity cost, especially when dealing with longer bit-length transmissions in the considered short block regime, despite providing enhanced or acceptable detection performance compared to a *conventional receiver*. Therefore, it is becoming realistic to consider alternative detection/decoding strategies that offer a favorable performance/complexity trade-off, compliant with such RM coded transmissions. Building on this challenge, we introduce the principle of block/segment encoding using *First-order RM* (FoRM) Codes, which are amenable to low-cost decoding through block-based *fast Hadamard transforms* (BFHT). The Block-based FHT has demonstrated to be cost-efficient with regard to decoding time, as it evolves from quadric to quasi-linear complexity with a manageable penalty in performance. Subsequently, incorporating an adaptive DMRS/data power adjustment, it becomes feasible to narrow the performance gap with respect to the conventional maximum likelihood receiver, leading to a good trade-off between performance and complexity to efficiently handle small payloads. Alongside this framework, in the search for alternative codes that could compete with or even outperform 3GPP RM codes, we are relying on the sparse vector coding (SVC) approach. The SVC scheme is simple, flexible, and of low complexity, making it suitable for short packet transmissions. The baseline decoding process is projection based, following a simplified Orthogonal Matching Pursuit. Hence, we propose an enhanced decoding approach leveraging LLR-based attention-weighted projection, which dynamically reweights received signal measurements according to their reliability, giving more importance to more reliable measurements. Simulation results over both Rayleigh flat fading and 3GPP TDL-C/TDL-D frequency-selective channels show that the proposed LLR-based attention-weighted projection decoding outperforms baseline projection decoding with both structured and random projection matrices. Moreover, we extend our analysis to a short-packet (3–11 bits) transmission framework over the 3GPP PUCCH. The results highlight that the proposed SVC scheme, employing a partial Hadamard spreading matrix combined with LLR-based attention-weighted projection decoding, can competitively rival 3GPP RM codes under optimal maximum likelihood (ML) decoding at very low BLER targets, while offering significant computational complexity advantages, making it a promising short-packet coding candidate for ultra-reliable, low-latency 6G short-block-length uplink/downlink 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.

Keywords : 5G NR, URLLC, Short data Transmission, Receiver design, Bit-Interleaved Coded Modulation metrics, 5G Polar codes, 5G LDPC codes, Unknown Channel State Information, Joint estimation and detection, Reed Muller code, Maximum likelihood decoding, Decoding via block-based Fast Hadamard Transform, Adaptive DMRS/Data power adjustment.

Résumé

Les communications ultra-fiables à faible latence dans le sillage de la 5G et au-delà suscitent une attention particulière au cours des dernières années. D'innombrables efforts sont entrepris pour améliorer les schémas de transmission, notamment dans la conception de récepteurs et d'autres algorithmes de communication en vue de répondre aux exigences de fiabilité et de latence. Cependant, bien que les formats de transmission de la 5G puissent permettre une transmission de paquets très courts grâce à l'utilisation de mini-slots, le rapport entre les informations d'entraînement et les données n'est pas nécessairement adapté à une transmission des paquets extrêmement courts. De plus, les formats de transmission de paquet court sont conçus avec des récepteurs quasi-cohérents conventionnel, qui peuvent être assez sous-optimaux dans de tels scénarios où l'estimation précise du canal est impossible en raison du caractère sporadique inherent de ces transmissions.
En termes simples, la conception de stratégies de transmission sophistiquées et innovantes dans le sillage des paquets courts revêt une importance capitale, contribuant à façonner la quatrième révolution industrielle (Industrie 4.0). Ainsi, cette thèse se concentre sur des aspects importants de la transmission de paquet courts, notamment la conception de récepteurs fiable et à faible complexité.
Nous nous intéressons, de prime abord, à la conception de récepteurs BICM avancés pour les paquets courts conformes aux transmissions codées polaires et LDPC. En effet, la pertinence du BICM devient particulièrement évidente dans les scénarios caractérisés par des canaux de communication sujets aux erreurs, nécessitant un niveau accru de fiabilité. Son efficacité dépend également des métriques de détection et de décodage sous-jacentes, d'où l'importance d'un équilibre nuancé entre performances et complexité est de mise. Les nouvelles métriques de réception proposées considèrent des scenarios dans lesquels les informations sur l'état du canal (CSI) sont de facto méconnues, visant à évaluer l'impact de différentes conditions de canal. Ainsi, nous présentons des récepteurs améliorés pour les courtes données dans la plage de 20 à 100 bits pour les scénarios de signalisation envisagés au-delà de la 5G/6G en évaluant leurs performances sur des canaux de blocs courts 5G. Nous examinons les métriques des récepteurs exploitant l'estimation et la détection conjointes (JED), qui conviennent aux situations où une faible densité de signaux de référence de démodulation (DMRS) est entrelacé avec des symboles de données codés et où les deux sont transmis sur un nombre restreint de symboles OFDM, de sorte qu'une estimation quasi parfaite du canal est impossible. Nous montrons qu'il est possible d'améliorer les performances et la sensibilité grâce à l'estimation-détection conjointe par blocs par rapport aux récepteurs BICM 5G OFDM classiques. Contrairement à la détection symbole par symbole conventionnel dans les systèmes BICM, où l'intervalle d'observation d'un bit codé particulier est limité au symbole dans lequel il est transmis, la méthode proposée effectue une détection conjointe par blocs sur une fenêtre glissante de symboles adjacents afin d'exploiter fondamentalement leurs dépendances statistiques. En conséquence, le LLR pour un bit codé particulier intègre les informations provenant de tous les symboles dans la fenêtre de détection, plutôt que d'être limité à son symbole hôte seul. L'évaluation des performances couvre les configurations SIMO et SU-MIMO, soulignant l'efficacité de la stratégie d'estimation-détection dans des scénarios réalistes de récepteurs. Nos résultats démontrent que lorsque les fenêtres de détection utilisées dans les unités métriques sont de l'ordre de quatre symboles modulés, les récepteurs proposés surpassent considérablement les récepteurs conventionnels et peuvent être utilisés pour obtenir des performances de détection proches de celles des récepteurs cohérents avec un CSI parfait. Parallèlement, nous examinons l'impact de la densité DMRS et la manière dont l'amplification de puissance DMRS peut être utilisée pour obtenir des performances supplémentaires.
En sus de cette contribution initiale, nous proposons aussi une conception de récepteurs améliorés et à faible complexité

pour des transmissions utilisant les codes Reed-Muller (RM). Récemment, les codes RM ont regagné un intérêt de recherche significatif en raison de leur similitude avec les codes polaires et de leur excellente performance de correction d'erreurs par l'entremise d'un décodage dit de maximum de vraisemblance (ML). De plus, les formes d'onde de la nouvelle radio (NR), telles que le canal de contrôle physique de liaison montante (PUCCH), ont été spécifiquement conçues pour transmettre de petites charges utiles avec des taux d'erreurs minimes dans des conditions exigeantes de rapport signal/bruit. La conception de récepteurs avancés capables de détecter, voire de décoder de manière fiable des courts paquets dans la plage de 3 à 11 bits avec une complexité moindre et une consommation d'énergie faible semble plus que nécessaire pour tirer pleinement parti de leur potentiel en matière d'URLLC et de leur utilisation dans les scénarios de signalisation NR. Ainsi, nous nous intéressons au schéma de transmission 3GPP PUCCH assisté par DMRS consistant en l'utilisation de codes Reed-Muller décodables via maximum de vraisemblance. En analysant la structure du récepteur, nous mettons en évidence que le terme non cohérent n'est généralement pas utilisé dans les récepteurs conventionnels, ce qui peut entraîner des pénalités susceptibles d'affecter de manière significative les performances du récepteur dans les régimes de fonctionnement des systèmes 5G/6G. En outre, dans les systèmes de communication actuels, la détection/décodage de paquets courts généralement se compose d'une estimation du canal par la méthode dite des moindres carrés suivie d'une détection quasi cohérente. Cependant, il convient de noter que dans les deux cas, les algorithmes sous-jacents sont suboptimaux en raison de la procédure de détection impliquant une estimation de canal séparée suivie d'une détection quasi cohérente. Parallèlement, l'usage d'un récepteur entièrement non cohérent entraînerait en conséquence un coût de complexité substantiel, notamment lorsqu'il s'agit de transmissions de longueur de bits plus longues dans le régime de blocs courts considéré. Partant de ce constat, nous introduisons le principe de codage par blocs ou segments en utilisant des codes RM de premier ordre (FoRM), qui se prêtent à un décodage peu coûteux grâce à des transformées d'Hadamard rapide par blocs (BFHT) ayant plus de mérite en termes de temps de décodage. Toutefois, cela s'accompagne de pénalités de performance gérables grâce à la méthode d'ajustement adaptatif de la puissance des séquences DMRS, ce qui permet, dans une certaine mesure, de réduire l'écart de performance par rapport aux récepteurs de vraisemblance maximale conventionnels. Parallèlement à ce cadre, dans notre recherche de codes alternatifs susceptibles de rivaliser avec les codes 3GPP RM , voire de les surpasser, nous nous appuyons sur l'approche du codage vectoriel clairsemé (SVC). Le schéma SVC est simple, flexible et peu complexe, ce qui le rend adapté aux transmissions de paquets courts. Le processus de décodage de base est basé sur la projection, suivant une poursuite orthogonale simplifiée. Par conséquent, nous proposons une approche de décodage par projection améliorée tirant parti de la projection pondérée par l'attention basée sur le LLR, qui repondère dynamiquement les mesures des signaux reçus en fonction de leur fiabilité, accordant plus d'importance aux mesures plus fiables pour la transmission de paquets courts codés par vecteur clairsemé. Les résultats de la simulation soulignent que le schéma SVC proposé, utilisant une matrice de projection d'Hadamard partielle combinée à un décodage par projection pondérée par l'attention, basé sur le LLR, peut efficacement concurrencer les codes 3GPP RM décodables via un décodage ML optimal, notamment dans des seuils de BLER très faibles, avec une complexité de décodage nettement inférieure à celle du décodage ML, ce qui en fait un candidat prometteur pour le codage de paquets courts sur les canaux UL/DL 6G à faible latence et ultra-fiables.

Mots clés : 5G NR, URLLC, Transmission de paquets courts, Conception de récepteurs, métriques de modulation codée, Codes polaires 5G, Codes LDPC 5G, Informations sur l'état du canal inconnu, Estimation et détection conjointes, Codes Reed-Muller, Décodage par maximum de vraisemblance, Décodage par blocs via la transformée d'Hadamard rapide, Ajustement adaptatif de la puissance DMRS/données.

List of Figures

List of Tables

List of Acronyms

5G Fifth Generation of Wireless Cellular Technology
6G Sixth Generation of Wireless Cellular Technology
BCH Broadcast Channel
BICM Bit-Interleaved Coded Modulation
BILCM Bit-Interleaved LDPC Coded Modulation
BIPCM Bit-Interleaved Polar Coded Modulation
BPSK Binary Phase Shift Keying
BLER Block Error Rate
BP Belief Propagation
CA-SCL CRC-aided Successive Cancellation List
CRC Cyclic Redundancy Check
CPS Cyber Physical System
CSI Channel State Information
DMRS Demodulation Reference Signal
DL-SCH Downlink Shared Channel
FHT Fast Hadamard Transform
FoRM First-order Reed Muller
IID Independent and Identically Distributed
HT Hadamard Transform
IIoT Industrial Internet of Things
ITU International Telecommunication Union
JED Joint Estimation and Detection
LBP Layered Belief Propagation
LDPC Low-Density Parity-Check
LLR Log-Likelihood Ratio
LOS Line-of-Sight
MAC Medium Access Control
M2M Machine To Machine
ML Maximum-Likelihood
MIMO Multiple-Input Multiple-Output
NLOS Non-Line-of-Sight
NR New Radio
OFDM Orthogonal Frequency Division Multiplexing
PCH Paging Channel
PHY Physical Layer
PxCCH Physical Uplink/Downlink Control Channel
PxSCH Physical Uplink/Downlink Shared Channel
QPSK Quadrature Phase Shift Keying
RM Reed Muller
SC Successive Cancellation
SCL Successive Cancellation List
SBL Short Block Length
SDT Short Data Transmission
SU Single User
TB Transport Block
TDL Tapped Delay Line
UL-SCH Uplink Shared Channel
URLLC Ultra-Reliable-Low-Latency Communication

List of Symbols and Notations

$\mathbf{x}$	:	Complex-valued vector.
$\mathbf{X}$	:	Complex-valued Matrix.
x	:	Complex/real-valued Scalar.
$\|\|\mathbf{x}\|\|$	:	Euclidean norm.
$\|\cdot\|$	:	The absolute value.
$\mathrm{tr}\{\cdot\}$	:	Trace of matrix.
$\|\cdot\|_{\mathsf{F}}$	:	Frobenius norm of matrix.
$\mathrm{I}_0(\cdot)$	:	Zero-th order modified Bessel function of the first kind.
$\mathrm{Re}(\cdot)$	:	Real part of a complex number.
$\mathbb{E}\{\cdot\}$	:	Statistical expectation.
$\mathbb{F}_2$	:	Galois field is denoted by $GF(2)$.
$\mathbf{I}$	:	Identity matrix with appropriate dimensions.
$\boldsymbol{\chi}_b^j$	:	The subset of constellation symbols $\{\mathbf{x}\}$ for which the $j-th$ bit of the corresponding binary label e is equal to $b = \{0, 1\}$.
$m \triangleq \log_2(\mathcal{M})$	:	The number of bits reqired to a symbol.
$\mathcal{M} \triangleq \|\boldsymbol{\chi}\|$	:	The cardinality of $\boldsymbol{\chi}$.
$\Lambda^j(\cdot)$	:	The log likelihood ratio, with $j = 1, 2, \ldots, m$.
†	:	The complex conjugate transpose or Hermitian.
$^\mathsf{T}$	:	The Matrix/vector transpose.
$\otimes$	:	The Kronecker product.
$\oplus$	:	The Binary addition.
$A \subseteq$B	:	Set A is a subset of Set B.
$\mathcal{O}(\cdot)$	:	The Big O complexity.
$\mathsf{d} = \biguplus_{b=1}^{B} \mathcal{D}_b$	:	The index sets $\mathcal{D}_b$ are pairwise disjoint and their union exactly covers d, i.e.,$\mathcal{D}_b \cap \mathcal{D}_{b'} = \emptyset$ for all $b \neq b'$, and $\bigcup_{b=1}^{B} \mathcal{D}_b = \mathsf{d}$.
$\mathscr{C}(N', K)$	:	A code with a codeword length of N' bits and K information bits.

Contents

Chapter 1

Introduction

> If you wish to accomplish a lasting work, be patient, be virtuous, be tolerant, and above all, be human.
>
> — Amadou Hampâté Bâ

1 Motivations

The emergence of 5G *new radio* (NR) marks a watershed moment in telecommunications, delivering highly scalable and flexible radio access technology essential to digital transformation across multiple sectors. This technical milestone has sparked a rush to look for applications in healthcare, industrial processes, transportation, and beyond. As a result, scholars and industry actors are being prompted to realign their approaches to transmission system design, necessitating future communication standards that consider critical parameters such as low latency, reduced energy consumption, increased connectivity, extended coverage, enhanced resilience, and improved capacity. Many significant reasons have contributed to 5G's increased prominence. Firstly, the exponential rise in mobile data traffic, especially driven by video streaming, necessitates a network capable of efficiently managing such volumes. Additionally, the proliferation of devices per user demands increased connectivity capabilities, and in some way the emergence of the Internet of Things adds to this challenge, since networks must handle billions of new devices. However, 5G goes beyond addressing these challenges; it also opens new perspectives by enabling ultra-low latency and high reliability use cases.

The ITU-R has delineated the primary usage scenarios for 5G in Recommendation ITU-R M.2083 for 2020 and beyond. These scenarios include enhanced *mobile broadband* (eMBB), catering to the need for increased data rates, high user density, and ample traffic capacity in hotspot and high-mobility scenarios. Additionally, *massive machine type communications* (mMTC) target the *Internet of Things* (IoT), emphasizing low power consumption and data rates for numerous connected devices. Lastly, *ultra-reliable and low latency communications* (URLLC) is intended for mission-critical applications that require high reliability and low latency. The interest of URLLC in 5G and beyond has prompted a focus on advanced transmission strategies, as underscored by various scholarly works. Emphasis is placed on the need for error rate performance and low complexity receivers to meet URLLC requirements. Interestingly, enhancements for ultra-reliable, low latency NR communications are envisaged within the Rel-16 standard, with enhanced *Industrial Internet of Things* (IIoTs) envisioned within the Rel-17 standard. Furthermore, specific applications in the fields of IIoT and *cyber-physical systems*(CPSs) entail the transmission of short packets, which, to some extent, pose fundamental limits that must be considered to meet strict latency and reliability constraints.

Likewise, the air interface of 6G is expected to build upon the 5G standard and address new paradigms for cyber-physical systems based on feedback and combining communications and sensing. Specifically, tight control loops will be required using the air interface to control 6G-compatible objects with high reliability, perhaps even demanding latency times lower than those achieved by current 5G technology, for instance, latency lower than 1 ms for the

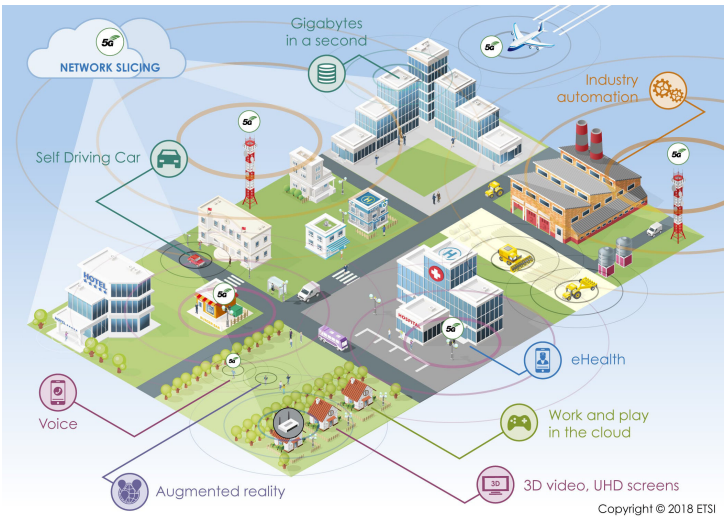


Figure 1.1: Why do we need 5G? [1].

uplink application layer in the microwave spectrum. Although 5G transmission formats may enable very short packet transmission through the use of mini-slots, the ratio between training information and data may not necessarily be suitable for extremely short data transmission. Since the more the demodulation reference signals (DMRSs) or pilots, the higher the transmission overhead. Conversely, few DMRSs result in lower channel estimation quality, therefore poor detection/decoding performance. There is some challenge to find a sweet spot in term of DMRS density.
Additionally, transmission formats are designed with conventional quasi-coherent receivers that may prove suboptimal in scenarios where precise channel estimation is impossible because of the sporadic nature of short packet transmissions. In other respects, at the demodulator level, conventional receivers use de facto *symbol-by-symbol* detection, which does not seem to perform well in such scenarios. One of the key applications lies in critical communications, such as those used by emergency services or in industrial control systems requiring extremely high levels of reliability and low latency. The same applies to control channels with feedback on channel state information, or for future combined sensing and communication paradigms requiring rapid sensory feedback to the network. Phrased directly, the design of enhanced physical-layer transmission strategies is of paramount relevance in beyond 5G and 6G URLLC contexts as well as for mission-critical applications supported by IIoTs and CPSs, contributing to shaping the *fourth industrial revolution* (Industry 4.0), where future factories will be subject to stringent requirements regarding throughput, latency, coverage, reliability, spectral efficiency, and minimal complexity.

2 Objectives

This thesis aims to provide insight into a number of key issues concerning short packet transmission, in particular the design of ultra-reliable low-latency receivers. The objectives are twofold, as listed below.

- **Enhanced BICM receiver design compliant with polar and LDPC coded transmissions.**

BICM stands as a prominent coding paradigm in wireless communication channels, serving as a cornerstone for contemporary high spectral efficiency systems and low spectral efficiency orthogonal modulation systems. Its significance 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 trade-off 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. There is a wealth of literature on BICM receivers from various perspectives, demonstrating their potential influence and relevance. Among the pioneers who sparked interest in BICM was the seminal work conducted by Caire *et al.* [88], wherein they provided a comprehensive analysis of BICM in terms of information rate and error probability, encompassing both coherent and non-coherent detection. We are particularly interested in BICM receivers for joint estimation and detection. This is particularly relevant as BICM OFDM emerges as an appealing prospect for future wireless networks, since OFDM reduces equalization complexity, and BICM provides reliable coded-modulations.
In addition, more reliable receivers can help to reduce errors in decoding user data or control information transmitted

via uplink or downlink channels (e.g. PxSCH, PxCCH), which can improve network stability and hence quality of service. Advanced receiver design seems relevant in regard to 5G, where efficient transmission is essential to ensure optimal network performance. In this way, we are particularly interested in the transmission of short packets. Short packet transmission is typically employed in 5G NR signaling scenarios. Short packets serve, to some extent, as vectors for embedding critical control information essential to network management and user connections. They find utility in various operations including device registration, session management, configuration of connection parameters, mobility management, and support for specific services like massive IoT communications or low-latency, high-throughput communication aspects, alongside other critical control functionalities. Figure 1.2 highlights some typical scenarios where short-packet transmissions are involved.

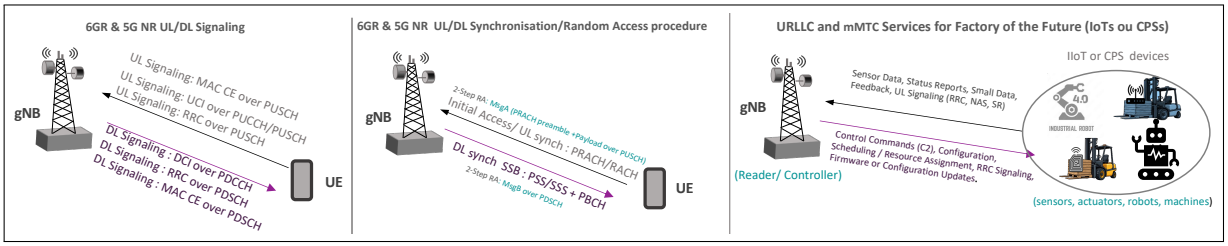


Figure 1.2: Short packet transmission: uscases and applications.

Therefore, we intend to explore the potential of advanced BICM receiver designs, compliant with 5G Polar and LDPC coded modulation transmissions, compared to conventional BICM receivers. Emphasis is set on enhancing their performance in handling short packet reception while approaching the performance levels of the ideal or coherent receiver.

- **Reliable low-complexity transceiver design compliant with Reed Muller coded transmissions.**

Recently, Reed-Muller (RM) codes have regained significant research interests due to their similarity with polar codes and their excellent error-correction performance via *maximum-likelihood* (ML) decoding. Furthermore, NR waveforms, including PUCCH, have been specifically designed to transmit small payloads with minimal error rates in challenging signal-to-noise conditions. To fully harness their URLLC potential and applicability to massive machine-type communications, robust coding techniques with low-complexity detection and decoding algorithms must be designed. Hence, there is a demand for improved receiver designs capable of accurately and reliably decoding short data transmissions ranging from 3 to 11 bits, while maintaining low computational complexity and power consumption. In this respect, we aim to explore the enhancement of short-block length detection/decoding strategies. For the primary case of interest here, we intend to investigate the possibility of enhancing standard 3GPP PUCCH transmission, assisted by *orthogonal demodulation reference signals* (DMRS) along with Reed-Muller codes. These 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 *estimator-correlator*, we 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 suboptimal because of the detection procedure involving separate channel estimation and *quasi-coherent detection*. Moreover, a *fully non-coherent receiver* also incurs a prohibitive complexity cost, particularly when dealing with longer bit length transmissions within the considered short block regime, 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, compliant with 5G Reed Muller coded transmissions.

3 Main contributions

The main contributions of this Thesis can be summarized in the following list.

- The **first contribution** consists of the design of novel BICM receiver metrics under imperfect CSI conditions, intending 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 LDPC coded formats. (a) We look into receiver metrics exploiting *joint estimation and detection* (JED), which are particularly amenable to configurations where low-density of *demodulation reference signals* (DMRS) are interleaved with coded data symbols. (b) We specifically address situations where accurate channel estimation is impossible, demonstrating that a well-conceived joint estimation–detection receiver, leveraging interleaved DMRS within the detection metric, can achieve remarkably significant performance gains over conventional 5G *orthogonal frequency division multiplexing* (OFDM) receivers, and can potentially approach the performance of a perfect CSI receiver, applicable to both uplink and downlink transmission scenarios.

Explicitly, our proposal consists of designing and utilizing novel soft-likelihood metrics that directly integrate channel estimation performed via joint least squares followed by averaging or smoothing across DMRS dimensions for bit-level LLR generation. Moreover, we apply to the underlying soft-likelihood metrics an advanced block-wise joint detection scheme defined over a detection window of four modulated symbols ($M = 4$). Unlike conventional symbol-by-symbol detection in BICM systems, where the observation interval for a given coded bit is confined to the symbol in which it is conveyed, the proposed block-wise detection approach fundamentally leverages the statistical dependencies between adjacent symbols; that is, the LLR for a given coded bit incorporates information from all symbols within the detection window, rather than being constrained to its host symbol alone, thereby enhancing detection reliability. Hence, our contributions span the following principal avenues. Initially, we introduce a BICM receiver metric specifically tailored for non-coherent fading channels in *single-input multiple-output* (SIMO) transmissions. These metrics effectively address challenges arising from both line-of-sight (LOS) and *non-line-of-sight* (NLOS) fading channels. Secondly, we extend the BICM receiver metric design to *single-user* MIMO systems, specifically addressing block fading channel conditions.

Meanwhile, we look into the impact of DMRS density and how DMRS power boosting may be used to achieve some extra performance.

Thus, these works have been communicated, respectively, in:

[1] **Mody Sy** and Raymond 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 6–8, 2023.

[2] **Mody Sy** and Raymond Knopp, "Receiver Metric Design for Short-Block Channels : A Perspective for Reliable 6G Signalling Scenarios", *2024 European Conference on Networks and Communications & 6G Summit* (EuCNC & 6G Summit), Antwerp, Belgium, June 3–6, 2024.

[3] **Mody Sy** and Raymond Knopp, "Enhanced BICM Receiver Metrics for Ultra-Reliable Low-Latency Short Block Channel Communications", *submitted to IEEE Transactions on Wireless Communications* (preprint is available on Arxiv: https://doi.org/10.48550/arXiv.2408.08660, August 16, 2024).

- The **second contribution** concerns of designing reliable low complexity transceivers for RM-coded transmissions and similar transmission schemes. Thus, we propose an approach to tackle the aforementioned challenge in RM-coded transmissions by introducing the principle of block (or segment) encoding using first-order RM (FoRM) codes, which is amenable to low-complexity decoding through block-based fast Hadamard transforms (FHT). The block-based FHT has demonstrated to be cost-efficient regarding decoding time, as it evolves from quadratic 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. Alongside this framework, in the search for alternative codes that could compete with or even outperform 3GPP RM codes, we are relying on the sparse vector coding (SVC) approach. The SVC scheme is simple, flexible, and of low complexity, making it suitable for short packet transmissions. The baseline decoding process is projection based, following a simplified Orthogonal Matching Pursuit. Although its performance in terms of BLER is comparable to 3GPP RM codes decoded by Maximum Likelihood, we propose an enhanced projection decoding approach leveraging LLR-based attention-weighted projection, which dynamically reweights received signal measurements according to

their reliability, giving more importance to more reliable measurements for sparse vector coded short packet transmsion. Simulation results over both Rayleigh flat fading and 3GPP TDL-C/TDL-D frequency-selective channels show that the proposed LLR-based attention-weighted projection decoding outperforms baseline projection decoding with both structured and random projection matrices. Moreover, we extend our analysis to a short-packet (3–11 bits) transmission framework over the 3GPP PUCCH. The results highlight that the proposed SVC scheme, employing a partial Hadamard spreading matrix combined with LLR-based attention-weighted projection decoding, can competitively rival 3GPP RM codes under optimal maximum likelihood (ML) decoding at very low BLER targets, while offering significant computational complexity advantages, making it a promising short-packet coding candidate for ultra-reliable, low-latency 6G short-block-length uplink/downlink channels.

These works were communicated in:

[4] **Mody Sy** and Raymond Knopp, "Enhanced Low-Complexity Receiver Design for Short Block Transmission Systems", *In the Proceedings of the 34th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications* (PIMRC 2023), Toronto, ON, Canada, Sept 5–8, 2023.

[5] **Mody Sy** and Raymond Knopp, "Low-complexity Block-Based Decoding Algorithms for Short Block Channels ", *In the Proceedings of the 16th IEEE AFRICON 2023*, Nairobi, Kenya, Sept 20–22, 2023.

[6] **Mody Sy**, "LLR-based attention-weighted projection decoding for sparse vector coded short-packet transmissions", *In IEEE International Conference on Communications (IEEE ICC 2026), Wireless Communications Symposiumn, Glasgow, Scotland, UK 24–28 May 2026.*

4 Outline

The remainder of this Thesis manuscript is divided into five main chapters.

- **Chapter** 2 presents the essential background for 5G polar and LDPC coded modulations, serving as a prelude to subsequent chapters, with a focus on their relevance to the 5G NR standard. We first highlight the design process underlying these channel codes in order to understand their foundations and fundamentals. In parallel, we provide a literature review to establish the state of knowledge in the field. Finally, we explore the modulation schemes commonly used in 5G NR communication channels and discuss their potential incorporation with 5G Polar and 5G LDPC coding for URLLCs.
- **Chapter** 3 stands as the inaugural part, detailing receiver design (part I). We start with the prior art in BICM receivers, stating the general framework of the polar coded/LDPC transmission procedure and the BICM receiver design that covers non-coherent fading channels in a SIMO transmission with the aim of efficiently addressing the challenges posed by LOS and NLOS channels, focusing on the performance of BICM detection in realistic base station receiver scenarios. We then present the numerical results to demonstrate the merits of the propositions outlined in the chapter.
- **Chapter** 4 is the subsequent part devoted to receiver design (Part II). Continuing from the preceding chapter, we initiate with the MIMO BICM receiver design, addressing SU-MIMO Rayleigh block fading channels and SU-MIMO LOS channels. Subsequently, we present the numerical results to show the merits of the propositions outlined in the chapter, considering both performance and complexity perspectives.
- **Chapter** 5 corresponds to the third part on receiver design (Part III). We address the design of a reliable low latency transceiver for Reed Muller coded transmission. First, we elucidate a fundamental limitation of conventional ML detection in the short-packet reception paradigm, highlighting the non-negligible performance loss caused by ignoring the so-called non-coherent energy term and how to overcome it to some extent. Secondly, we then propose a design framework based on block/segment encoding and decoding within RM coded transmission, which leads to the proposal of a low-complexity receiver with *block-based fast Hadamard* transforms coupled with an adaptive DMRS power adjustment strategy to approach the performance of conventional ML receivers. Finally, in the third part, we present a LLR-Based Attention-weighted projection decoding for sparse vector-coded short packet transmissions. We then present the numerical results to show the merits of the proposals presented in this chapter in terms of performance and complexity.

- **Chapter** 6 concludes the manuscript, providing an insight into the primary contributions of this research and opening perspectives for future work.

Chapter 2

5G NR Coded Modulation Fundamentals

Contents

1 Channel Coding Fundamentals

Channel coding has appeared since Shannon applied probability theory to study the communication system. Shannon [2] shows that the errors induced by the noisy channel can be reduced to the desired level by using a proper coding scheme for a given transmission rate less than or equal to channel capacity.

In order to optimize the performance of communication channels, different techniques can be implemented to reduce the effect of noise, disturbances or distortions and to increase the channel throughput until it approaches its maximum capacity. It is possible to physically modify the signal or change its content through modulation and channel coding, respectively.

In 5G NR, data and control streams from PHY to MAC layer are encoded /decoded to offer transport and control services over the radio transmission link. Channel coding scheme is a combination of error detection, error correcting, rate matching, interleaving and transport channel or control information mapping onto/splitting from physical channels. Table 2.1 and Table 2.2 show the usage of coding scheme for the different types of Transport channels and for the different control information types [132].

Table 2.1: Usage of channel coding scheme for *transport channels* (TrCHs)[132].

TrCH	**Coding scheme**
LDPC	UL-SCH DL-SCH
Polar code	PCH BCH

Table 2.2: Usage of channel coding schemes within 5G NR UL/DL control channels [132].

Control Information	Coding scheme
DCI	Polar code Block code
UCI	Polar code

1.1 Linear Block Codes:

Most error-correcting codes are linear codes. These linear codes can be represented systematically, so they are systematic codes. Linear block codes can be understood in a variety of ways, and some of the most commonly used in many contexts are *Reed Solomon* (RS) codes [5], Hamming codes [3], CRC codes, *Bose–Chaudhuri–Hocquenghem* (BCH) codes [6] and LDPC codes [10]. It is possible to represent a linear block code in matrix form or in the form of a graph factor.

1. *Generator Matrix:* $\mathbf{G}$ is the generator matrix of dimension $\mathsf{K} \times \mathsf{N}$. The matrix is said to be *systematic*: If $\mathbf{G} = [\mathbf{P}\ \mathbf{I}_{\mathsf{K}}]$ where $\mathbf{I}_{\mathsf{K}}$ and $\mathbf{P}$ is called the parity matrix of dimension $(\mathsf{K} \times \mathsf{N} - \mathsf{K})$. In this case, the code is formed by parity bits followed by message bits.
   The matrix $\mathbf{G}$ is said to be *non-systematic*: if $\mathbf{G} = [\mathbf{P}\ \mathbf{M}]$ where $\mathbf{M}$ is any matrix of size $(\mathsf{K} \times \mathsf{K})$.
2. *Parity Check Matrix:* $\mathbf{H}$ is called the parity check matrix denoted as $\mathbf{H} = [\mathbf{I}_{\mathsf{N}-\mathsf{K}}\ \mathbf{P}^{\mathsf{T}}]$, where $\mathbf{P}^{\mathsf{T}}$ is the transpose of the parity matrix $\mathbf{P}$. Thus, it can be easily shown that $\mathbf{G} \cdot \mathbf{H}^{\mathsf{T}} = \mathbf{0}$, where operations are performed in $\mathbb{F}_2$.
3. *Syndrome Checking:* For each code word $\mathbf{c}$: $\mathbf{c} \cdot \mathbf{H}^{\mathsf{T}} = \mathbf{m} \cdot \mathbf{G} \cdot \mathbf{H}^{\mathsf{T}} = \mathbf{0}$, where $\mathbf{m} = [\mathrm{m}(0), \mathrm{m}(1), \ldots, \mathrm{m}(\mathsf{K}-1)]$. If $\mathbf{c} \cdot \mathbf{H}^{\mathsf{T}} \neq \mathbf{0}$, it means there are errors in the received bits.

1.2 5G NR LDPC Codes

It is a curious twist of history that LDPC codes should have been largely unnoticed for so long. LDPC codes were originally proposed in 1962 by Gallager [10] . At that time, the codes might have been overlooked because contemporary investigations in concatenated coding overshadowed LDPC codes and because the hardware of the time could not support effective decoder implementations[85]. They therefore remained discrete until 1996 after the introduction of iterative decoding, initiated by the turbo codes[7]. Since then, LDPC codes have shown interesting performance and a relatively uncomplicated implementation. MacKay, working on Turbo codes at that time, gave a second birth to LDPC codes [8] and brings LDPC codes back into fashion. This article by Mackay presents constructions of LDPC codes and shows their good performance. Later, Luby, introduces irregular LDPC codes [9] characterized by a parity check matrix for which the distribution of the number of non-zero elements per row and/or column is not uniform. LDPC codes are linear block codes based on sparse parity-check matrix. It is forgotten for dozens of years because of the limited computation ability. In recent years, LDPC codes attract more attention because of their efficient decoding algorithms, excellent error-correcting capability, and their performance close to the Shannon limit for large code lengths. LDPC coding is currently adopted in 5G NR for both uplink and downlink shared transport channels. Given that 5G must support high data rates of up to 20 Gbps and a wide range of block sizes with different coding rates for data channels and *hybrid automatic repeat request* (HARQ), LDPC codes are a de facto candidate to meet these requirements. Indeed, the base graphs defined in 3GPP TS 38.212 [132] are structured parity-check matrix, which can efficiently support HARQ and rate compatibility that can support arbitrary amount of transmitted information bits with variable code rates. While Polar codes are applied to 5G NR control channels, LDPC codes are suitable for 5G NR shared channels due to its high throughput, low latency, low decoding complexity and rate compatibility. Another advantage of the NR 5G codes is that the performance of the LDPC codes has an error floor around or below the 10^5 BLER for all code width and code rates, making LDPC codes, an essential channel coding scheme for URLLC application scenarios.

1.2.1 State-of-art LDPC codes

In reviewing the literature, significant efforts have been directed towards enhancing the error correction performance of 5G communication systems. The discourse surrounding channel coding in cellular systems, particularly for 5G, initially emphasized turbo, LDPC, and polar codes, with LDPC codes adopted for eMBB data and polar codes for control [19]. In this regard, Srisupha *et al.* [20] developed an experimental kit demonstrating LDPC encoding processes, emphasizing flexibility between software and hardware LDPC encoders. Subsequently, Belhadj *et al.* [21]

compared error correction performance between LDPC and polar codes in 5G *machine-to-machine communications*, highlighting specific requirements for different M2M applications. Later on, Li *et al.* [22] proposed dynamic scheduling strategies to reduce decoding complexity and improve error correction performance for short LDPC codes. Next, in a perspective of accelerated LDPC decoding, Tian and Wang [23] presented a base graph-based static scheduling method for layered decoding of 5G LDPC codes, achieving notable reductions in iteration numbers and performance enhancements. In [14, 15], authors investigated the structure and features of the base graphs [132], showing that the usage of a circularly shifted identity matrix known as the permutation matrix can greatly reduce the memory requirement for implementation. Indeed, 5G LDPC base graphs design aims to provide row orthogonality for fast and reliable decoding. Row orthogonality in 5G LDPC base graphs design can somehow reduce decoding latency. Both base graphs and all code rates involve puncturing code bits associated with the first two circulant columns prior to transmission, targeting high-weight columns for performance enhancement. Therefore, puncturing serves as a means of improving overall system performance [15]. Additionally, 5G LDPC fully utilizes the double diagonal structure of the base graphs. Due to the characteristic of base graphs, the double-diagonal structure can make LDPC encoding more efficient. Conversely, [11] proposed a novel efficient encoding method and a high-throughput low-complexity encoder architecture for 5G NR LDPC.

Furthermore, efforts have been directed towards optimizing implementations of these codes for heightened performance on software and hardware targets. For instance, Liao *et al.* [29] present a high-throughput LDPC encoding on a single *graphics processing unit* (GPU), while Tarver *et al.* [30] explore GPU-based LDPC decoding, showcasing potential for high throughput and low latency applications in 5G and beyond. Additionally, hardware architectures have been developed to efficiently decode LDPC codes, as seen in the work of Nadal and Baghdadi [31], who propose a highly parallel *field-programmable gate array* (FPGA) architecture. Conversely, Xu *et al.* [32] focus on software decoding with SIMD acceleration on Intel Xeon CPUs, achieving notable throughput with low latency. Sy [33] proposed optimisation strategies for low-latency 5G LDPC decoding over GPPs. Meanwhile, Li *et al.* [34] addressed the challenge of achieving high throughput rates, proposing a multicore LDPC decoder architecture achieving up to 1 Tb/s throughput for beyond 5G systems, while Aronov *et al.* [35] compare LDPC decoding performance between GPU and FPGA platforms, stressing the need for further optimization, particularly in reducing latency for GPU-based solutions.

Moreover, the selection of coding schemes for 5G eMBB services has been a focus of interest, with *quasi-cyclic* LDPC and polar codes chosen for data and control channels, respectively. Rao and Babu. [24] highlighted the importance of these codes, particularly QC-LDPC, and Wu *et al.* [25] proposed an efficient QC-LDPC implementation for 5G NR, enhancing throughput by matrix pruning. Ivanov *et al.* [27] introduced a novel concatenated code construction comprising outer and inner LDPC codes, demonstrating reduced decoding complexity and superior performance. Additionally, Song *et al.* [28] emphasized the significance of well-designed LDPC codes, particularly QC-LDPC codes, in approaching channel capacity and enabling high-speed data transmission. In parallel, LDPC codes' adoption in 5G standards underscores their importance in broadcasting and cellular communication systems Ahn *et al.* [36]. Moreover, Cui *et al.* [37] tackled the challenge of designing high-performance and area-efficient decoders, while Trung *et al.* [38] and Jayawickrama and He [39] proposed adaptations and improved algorithms for LDPC decoding in 5G NR. Additionally, Li *et al.* [40] focused on LDPC code design for specific 5G scenarios, proposing optimization methods and an improved decoding algorithm. Wu and Wang [41] provided insights into LDPC decoding latency in 5G NR, guiding decoder design to meet high throughput and low latency requirements. In Sun and Jiang. [26] proposed a hybrid decoding algorithm for LDPC codes in 5G, where normalized min-sum algorithm (NMSA) decoding and linear approximation are combined, with only a slight increase in complexity for NMSA and an improved performance much closer to *belief propagation* BP decoding, especially for low-rate codes.

Recently, there has been a growing interest in applying deep learning techniques to various aspects of 5G communications, as reviewed by Dai *et al.* [42]. Shah and Vasavada [43] proposed *normalized least-mean-square* (NLMS) algorithms to enhance the decoding performance of 5G LDPC codes, leveraging *deep neural networks* (DNNs) to optimize parameters. Tang *et al.* [44] introduced a scheme combining model-driven deep learning with traditional BP decoding algorithms to adapt LDPC codes for different 5G scenarios. Lastly, Andreev *et al.* [45] investigated the application of DNNs to improve the decoding algorithms of short QC-LDPC codes in the 5G standard, addressing the curse of dimensionality problem and enhancing performance.

1.2.2 Foundations and Fundamentals

The LDPC code is presented in matrix form $[\mathbf{G}, \mathbf{H}]$, where $\mathbf{G}$ is the generator matrix of the code and $\mathbf{H}$ is the *parity check matrix* (PCM). The parity matrix $\mathbf{H}$ is sparse, containing very few ones. It can be represented in the

form of a *Tanner graph*. This graph consists of two types of nodes: *bit nodes* (BNs) and *check nodes* (CNs) which are connected by edges. The BNs and CNs correspond, respectively, to the columns (code word bits) and rows (parity constraints) of the matrix $\mathbf{H}$. A variable node i is connected to a check node j if $\mathbf{H}(i;j) = 1$. This graph is bipartite, since nodes of the same type cannot be connected (i.e., a CN cannot be connected to another CN). Tanner graphs are commonly used to represent the parity matrix in LDPC codes. A Tanner graph leads to decoding algorithms of fairly low complexity [85].

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{2.1}$$

The number of variable nodes equals the number of received bits (N), which is equivalent to the number of columns in the matrix $\mathbf{H}$. Similarly, the number of parity check nodes corresponds to the number of rows $(N - K)$ in the matrix $\mathbf{H}$.

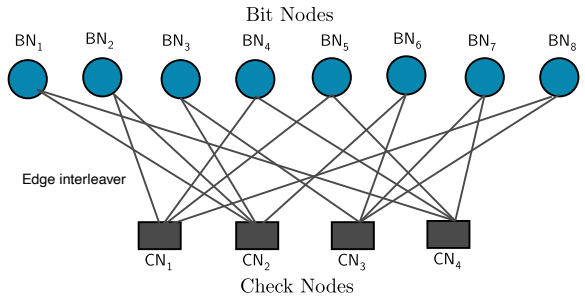


Figure 2.1: Parity check matrix $\mathbf{H}(4, 8)$ and the corresponding Tanner graph

The most important element in the realization of a good LDPC code is the $\mathbf{H}$ matrix which will condition the quality of the iterative decoding. There are four important parameters to respect in order to obtain a good $\mathbf{H}$ parity check matrix.

1. *Code Rate*. We can increase the rate of an LDPC code by adding 1s in the lines of the Matrix $\mathbf{H}$.
   If the parity matrix is regular, it can also be denoted (N, w_c, w_d), where w_b represents the weight of a row and w_b represents the weight of a column. The yield can be calculated as a function of w_c and w_b by $\mathrm{R} = 1 - w_b/w_c$
2. *short cycles (girth)*. The generation of short cycles in the equivalent Tanner graph of the code must be avoided. A cycle represents from a given variable node the set of parity and variable nodes that will be connected to it until we fall back on the starting variable node. We call girth the minimum cycle length that can be encountered in a Tanner graph. Short cycles are very penalizing because they involve few intermediate nodes, and thus the extrinsic information they generate during decoding becomes quickly and strongly correlated.
3. *Size of* $\mathbf{H}$. A large matrix allows better coding Rate and consequently better performance.
4. *Construction algorithm of* $\mathbf{H}$. To construct the regular parity matrix, there are mainly two methods: the random method and the deterministic method. The most known random methods in the literature for the construction of the $\mathbf{H}$-matrix are the Gallager method [10] and the *progressive Edge Growth* (PEG) method [12] which makes it possible to create graphs with large girths. For the deterministic method, the most used is the Quasi-cyclic method. The latter uses a deterministic construction based on a circular permutation of the identity matrix [13, 25]. QC-LDPC codes belong to the class of structured codes that are relatively easier to implement without significantly compromising the performance of the code. Well-designed QC-LDPC codes have been shown to outperform computer-generated random LDPC codes, in terms of bit-error rate and block-error rate performance and the error floor. These codes also offer merits in decoder hardware implementation due to their cyclic symmetry, which results in simple regular interconnection and modular structure[18].

1.2.3 3GPP 5G LDPC Codes

NR LDPC code is a family of QC-LDPC codes. It is constructed from a matrix named $\mathbf{H}_{\mathrm{BG}}$ of dimension $M \times N$ called *base graph* matrix $\mathbf{BG}$. The $\mathbf{H}_{\mathrm{BG}}$ matrices are selected in the 5G NR coding process according to the coding

rate and the length of the transport block or code block. Thus, for $\mathbf{BG}_1$ ($N = 68$, $M = 46$) and for $\mathbf{BG}_2$ ($N = 52$, $M = 42$). Since $\mathbf{BG}_1$ is targeted for larger block length $\mathsf{K} \leq 8448$ and coding rates between $1/3 \leq \mathrm{R} \leq 8/9$, $\mathbf{BG}_2$ is employed for small blocks $\mathsf{K} \leq 3840$ and coding rates between $1/5 \leq \mathrm{R} \leq 2/3$.

For $\mathbf{BG}_1$, $\mathrm{K} = 22\mathrm{Z_c}$ and for $\mathbf{BG}_2$, $\mathrm{K} = 10\mathrm{Z_c}$, where K is the maximum number of information bits, and $\mathrm{Z_c}$ is the lifting size shown in Table 2.3. There are 51 lifting sizes from 2 to 384 for each base graph. Both $\mathbf{BG}_1$ and $\mathbf{BG}_2$ have the same block structure. The columns include information columns, core parity columns, and extension parity columns. The rows are divided into core check rows and extension check rows.

For $\mathbf{BG}_1$, $\mathbf{A}$ is a 4×22 matrix; $\mathbf{E}$ is a 4×4 matrix; $\mathbf{O}$ is 4×42 all zero matrix; $\mathbf{B}$ is a 42×22 matrix; $\mathbf{C}$ is a 42×4 matrix; $\mathbf{I}$ is 42×42 identity matrix;
For $\mathbf{BG}_2$, $\mathbf{A}$ is a 4×10 matrix; $\mathbf{E}$ is a 4×4 matrix; $\mathbf{O}$ is 4×38 all zero matrix; $\mathbf{B}$ is a 38×10 matrix; $\mathbf{C}$ is a 38×4 matrix; $\mathbf{I}$ is 38×38 identity matrix; Sub-matrix $\mathbf{E}$ is a double diagonal matrix that is benefit for encoding. An example of $\mathbf{BG}_1$ with set *index of listing size* (iLS) $=1$ in 3GPP TS 38.212[132] standard is shown in Figure 2.2. In order to distinguish with the number 1 in base graphs in 3GPP TS 38.212 [132] standard, null value in the base graph will be replaced by -1. In 3GPP TS 38.212 [132] standard, the maximum lifting size value for each set of iLS is shown in

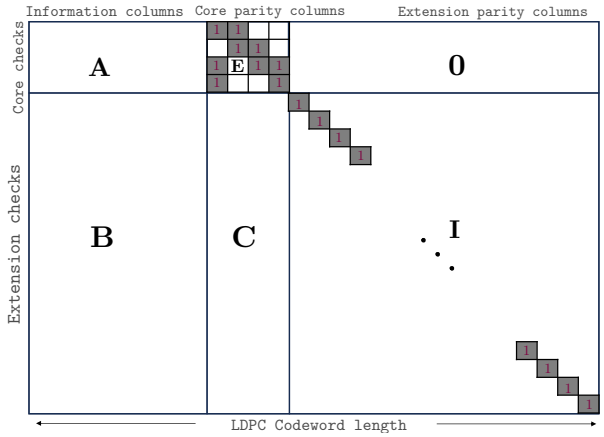


Figure 2.2: 3GPP NR Base graphs structure.

Table 2.3. The value of each element $\mathsf{p}_{\mathsf{i},\mathsf{j}}$ also known as the circular shift value is from -1 to 383, which is a property of the $\mathbf{BG}$s.

Table 2.3: Sets of LDPC lifting size [132]

Set index (iLS)	Set of lifting sizes (Z)
0	2, 4, 8, 16, 32, 64, 128, 256
1	3, 6, 12, 24, 48, 96, 192, 384
2	5, 10, 20, 40, 80, 160, 320
3	7, 14, 28, 56, 112, 224
4	9, 18, 36, 72, 144, 288
5	11, 22, 44, 88, 176, 352
6	13, 26, 52, 104, 208
7	15, 30, 60, 120, 240

NR LDPC codes also offer an additional coding advantage at lower code rates, rendering them suitable for scenarios requiring high reliability. Regarding decoding complexity, opting for $\mathbf{BG}_2$ proves advantageous due to its compactness and utilization of a larger lifting size element, translating to enhanced parallelism compared to $\mathbf{BG}_1$. The decoding latency tends to correlate with the number of non-zero elements in the base graph, with $\mathbf{BG}_2$ exhibiting significantly lower latency than $\mathbf{BG}_1$ for a given code rate, owing to its fewer non-zero elements [18].
Furthermore, the parity-check matrix $\mathbf{H}$ is obtained by replacing each element of the base graph $\mathbf{H}_{\mathrm{BG}}$ with a $\mathrm{Z_c} \times \mathrm{Z_c}$ matrix, according to the following rules.

- Each element of value -1 in $\mathbf{H}_{\mathrm{BG}}$ is replaced by a null matrix of size $\mathrm{Z_c} \times \mathrm{Z_c}$.
- Each element of value 0 in $\mathbf{H}_{\mathrm{BG}}$ is replaced by an identity matrix $\mathbf{I}$ of size $\mathrm{Z_c} \times \mathrm{Z_c}$.

- Each element of value from 1 to $Z_c - 1$ in $\mathbf{H}_{BG}$ which is denoted by $p_{i,j}$ is replaced by a circular permutation matrix $\mathbf{I}(p_{i,j})$ of size $Z_c \times Z_c$, where i and j are the row and column indices of the element, and $\mathbf{I}(p_{i,j})$ is obtained by circularly shifting the identity matrix $\mathbf{I}$ of size $Z_c \times Z_c$ to the right $p_{i,j}$ times [134].

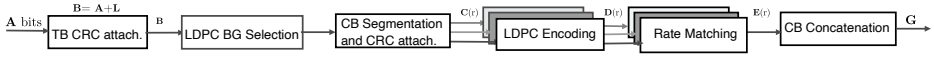
The main advantage of using a circularly shifting identity matrix is that it can reduce the memory requirement for implementation [15].
To simplify, a small example can be used to explain the principle of obtaining the parity check matrix $\mathbf{H}$. Hence, assuming that $\mathbf{H}_{BG}$ is a given base graph matrix with lifting size Z_c=4,

$$\mathbf{H}_{BG} = \begin{bmatrix} 2 & -1 & 1 & 3 & 0 & -1 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ -1 & 3 & 2 & 1 & -1 & 0 \end{bmatrix}, \tag{2.2}$$

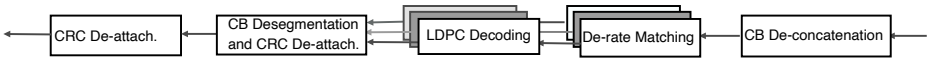
the corresponding parity check matrix $\mathbf{H}$ is shown to be:

$$\mathbf{H}= \begin{bmatrix} 0&0&1&0&0&0&0&0&0&1&0&0&0&0&0&1&1&0&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0&0&1&0&1&0&0&0&0&1&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&1&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&1&0&0&0&0 \\ 0&1&0&0&1&0&0&0&0&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0 \\ 0&0&1&0&0&1&0&0&0&0&0&0&0&1&0&0&0&1&0&0&0&1&0&0 \\ 0&0&0&1&0&0&1&0&0&0&0&0&0&0&1&0&0&0&1&0&0&0&1&0 \\ 1&0&0&0&0&0&0&1&0&0&0&0&0&0&0&1&0&0&0&1&0&0&0&1 \\ 0&0&0&0&0&0&0&1&0&0&1&0&0&1&0&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0&0&1&0&0&1&0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&1&0&0&1&0&0&0&0&0&0&1&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&1&0&0&1&0&0&1&0&0&0&0&0&0&0&0&0&0&1 \end{bmatrix}. \tag{2.3}$$

The overall LDPC transmission chain from MAC/PHY layer processing schematic is depicted in Figure 2.3, which describes the transmit-end for the PUSCH/PDSCH supporting a transport block CRC attachment, LDPC base graph selection, code block segmentation and code block CRC attachment, LDPC encoding, rate matching code block concatenation. The receiver end is therefore the transmitter end in the reverse flow. At the transmitter end, we can list



(a) Transmitter end.



(b) Receiver end.

Figure 2.3: Conceptual illustration of 5G LDPC transceiver chain.

the following components.

1. *Transport block CRC attachment*.
   CRC is an error detection code used to measure BLER after decoding. The entire transport block is used to calculate CRC parity bits. Assume that the transport message before CRC attachment is $a(0), a(1), \ldots, a(A-1)$, where A is the size of the transport block message. Parity bits are $p(0), p(1), \ldots, p(L-1)$, where L is the number of parity bits. The parity bits are generated by one of the following cyclic generator polynomials[132]. if $A > 3824$, the generator polynomial $g_{CRC24A}(\beta)$ is used.

$$g_{CRC24A}(\beta) = \left[\beta^{24} + \beta^{23} + \beta^{18} + \beta^{17} + \beta^{14} + \beta^{11} + \beta^{10} + \beta^{7} + \beta^{6} + \beta^{5} + \beta^{4} + \beta^{3} + \beta + 1\right]. \tag{2.4}$$

   The length of parity bits $L = 24$. Otherwise, the generator polynomial $\mathbf{g}_{CRC16}(\beta)$ is used.

$$g_{CRC16}(\beta) = \left[\beta^{16} + \beta^{12} + \beta^{5} + 1\right]. \tag{2.5}$$

   The length of parity bits $L = 16$. The message bits after attaching CRC are $b(1), b(2), \ldots, b(B)$, B represents the transport block information size with CRC bits such that $B = A + L$.

$$b_k = \begin{cases} a_k, & \texttt{for } k = 0, 1, \ldots, A-1 \\ p_{k-A}, & \texttt{for } k = A, A+1, \ldots, A+L-1. \end{cases} \tag{2.6}$$

The CRC value was determined to satisfy the probability of misdetection of the TB with BLER $\sim 10^6$ as well as the inherent error detection of LDPC code.

2. *LDPC base graph selection*.
LDPC $\mathbf{BG}$ is selected based on the transport block message size A and transport block coding rate R. If $A \leq 292$, or if $A \leq 3824$ and $\mathrm{R} \leq 0.67$, or if $\mathrm{R} \leq 0.25$, LDPC $\mathbf{BG}_2$ is used. Otherwise, LDPC $\mathbf{BG}_1$ is used [132].

3. Code block segmentation and code block CRC attachment.
The input message to CB segmentation is a transport message with CRC, denoted as $b(1), b(2), \ldots, b(B)$, where B is the input message length. Assume that the maximum code block length is K_{cb}, where $\mathsf{K}_{cb} = 8448$ for $\mathbf{BG}_1$ and $\mathsf{K}_{cb} = 3840$ for $\mathbf{BG}_2$. Code block segmentation is based on the following rules.
Assume that C is the number of code blocks.
`if` $B \leq \mathsf{K}_{cb}$,

$$C = 1, \quad L = 0, \quad B_r = B. \tag{2.7}$$

`Otherwise`,

$$C = \left[B / \left(\mathsf{K}_{cb} - L\right)\right], \quad L = 24, \quad B_r = B + C \cdot L. \tag{2.8}$$

Assume that the output of code block segmentation is $c_r(0), c_r(1), \ldots, c_r(\mathrm{K}_r - 1)$, $\mathrm{K}_r = \mathrm{K}$ is the number of bits for the r-th code block. For $\mathbf{BG}_1$, $\mathrm{K} = 22\mathrm{Z_c}$ and for $\mathbf{BG}_2$, $\mathrm{K} = 10\mathrm{Z_c}$, where $\mathrm{Z_c}$ is a lifting size that is the minimum value of $\mathrm{Z_c}$ in all sets of lifting sizes in Table 2.3 which can meet formula (2.9).

$$\mathsf{K}_b \cdot \mathrm{Z_c} \geq \mathrm{K}_r \tag{2.9}$$

Where K_r is the number of information and CRC bits in a code block and $\mathrm{K}_r = B_r / C$. K_b is related with LDPC base graph type and the size of input message B, shown in Table 2.4.

Table 2.4: K_b value [[132]

BG	B	K_b
1	all	22
2	$B > 640$	10
2	$560 < B \leq 640$	9
2	$192 < B \leq 560$	8
2	$B \leq 192$	6

The output of code block segmentation $c_{r,k}$ is calculated as following,
If $C = 1$,

$$c_{r,k} = \begin{cases} b_k, & \texttt{for } 1 \leq k \leq B \\ \texttt{NULL}, & \texttt{for } B+1 \leq k \leq \mathrm{K}. \end{cases} \tag{2.10}$$

If $C > 1$, block code should be attached CRC using the generator polynomial $\mathrm{g_{CRC24B}}(\beta)$, the length of parity bits $L = 24$.

$$\mathrm{g_{CRC24B}}(\beta) = \left[\beta^{24} + \beta^{23} + \beta^6 + \beta + 1\right]. \tag{2.11}$$

Assume that CRC parity bits are $p_r(1), p_r(2), \ldots, p_r(L)$,

$$c_{r,k} = \begin{cases} b_k, & \texttt{for } 1 \leq k \leq \mathrm{K}_r - \mathrm{L} \\ p_{r(k+L-\mathrm{K}_r)}, & \texttt{for } \mathrm{K}_r - \mathrm{L} + 1 \leq k \leq \mathrm{K}_r \\ \texttt{NULL}, & \texttt{for } \mathrm{K}_r + 1 \leq k \leq \mathrm{K}, \end{cases} \tag{2.12}$$

where $1 \leq r \leq C$ and K is the maximum number of information bits for base graphs.

4. *LDPC encoding*.
Each CB message is encoded independently. The input bit sequence in a CB to be passed to the LDPC encoder can be represented as $\mathbf{c}_r = [c_r(0), c_r(1), \ldots, c_r(\mathrm{K}_r - 1)]^{\mathsf{T}}$, where K_r is the number of information bits within a CB to

encode, the redundant bits are called parity bits denoted by $\mathbf{w} = [w(0), w(1), \ldots, w(\mathrm{N}_r + 2\mathrm{Z_c} - \mathrm{K}_r - 1)]^{\mathsf{T}}$. The output LDPC coded bits are denoted by $d_r(0), d_r(1), \ldots, d_r(\mathrm{N}_r - 1)$ where $\mathrm{N}_r = 66\mathrm{Z_c}$ for $\mathbf{BG}_1$ and $\mathrm{N}_r = 50\mathrm{Z_c}$ for $\mathbf{BG}_2$, where the value of lifting factor $\mathrm{Z_c}$ is given in Table 2.3. The LDPC encoding is based on the following procedure [132].

(a) Find the set with index iLS in Table 2.3 which contains $\mathrm{Z_c}$.

(b) Set $d_{r,k-2\mathrm{Z_c}} = c_k, \forall k = 2\mathrm{Z_c}, \ldots, \mathrm{K}_r - 1$.

(c) Generate $\mathrm{N}_r + 2\mathrm{Z_c} - \mathrm{K}_r$ parity bits $\mathbf{w} = [w(0), w(1), \ldots, w(\mathrm{N}_r + 2\mathrm{Z_c} - \mathrm{K}_r - 1)]^{\mathsf{T}}$ such that $\mathbf{H} \times [\mathbf{c}_r \ \mathbf{w}]^{\mathsf{T}} = \mathbf{0}$.

(d) The encoding is performed in $\mathbb{F}_2$.

(e) Set $d_{r,k-2\mathrm{Z_c}} = w_{k-\mathrm{K_r}}, \forall k = \mathrm{K_r}, \ldots, \mathrm{N}_r + 2\mathrm{Z_c} - 1$.

5. *Rate matching*:
The rate matching aims to adapt different code rates. Rate matching is based on *redundancy version* (RV) from 0 to 3 [132]. Each RV divides the base graph, excluding the first two columns, into four chunks at different positions. Note that the first two columns are always punctured to improve performance. RV 0 is well suited for the first transmission and has good self-decodability [49].
Hence, the rate matching is carried out on each code block independently. Assume that the coded message bit output from the LDPC encoder of the $r - th$ code block is $d_r(1), d_r(2), \ldots, d_r(\mathrm{N}_r)$ and E_r is the length of the output message after performing rate matching on this $r - th$ coded message. Thus, the output bit-message from the rate matcher is denoted by $e_r(1), e_r(2), \ldots, e_r(E_r)$ which is calculated using the following equation:

$$e_{r,k} = d_{r,k}, \ \texttt{if} \ d_{r,k} \neq \texttt{NULL}, \ \texttt{where} \ 1 \leq k \leq E_r, \quad 1 \leq r \leq C. \tag{2.13}$$

6. *Code block concatenation*:
The CB concatenation aims to concatenate all code blocks message to a sequence of transport block message, which will be transmitted through the physical channel. Assume that the output message of code block concatenation is $g(1), g(2), \ldots, g(G)$, where G is the desired length of the message of the transport block.

$$g_\ell = e_{r,k}, \ \texttt{where} \ 1 \leq \ell \leq G, \quad 1 \leq r \leq C, \quad 1 \leq k \leq E_r. \tag{2.14}$$

Conversely, the receiver counterpart, is simply the reverse flow of the transmitter, and set out as follows :

1. *Code block de-concatenation.*
The CB de-concatenation is used to break the transport block message into C numbers of code blocks message. Assume that the input message to code block de-concatenation is $y(0), y(1), \ldots, y(G-1)$. The output message from code block de-concatenation is $f_r(0), f_r(1), \ldots, f_r(E_r - 1)$.

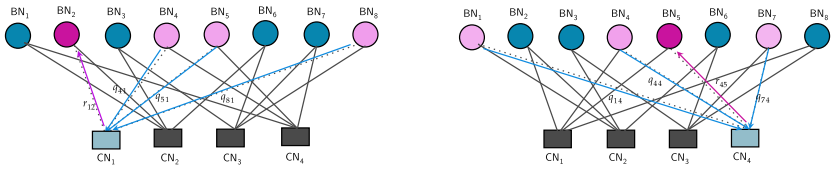
$$f_{r,k} = y_\ell, \ \texttt{where} \ 1 \leq \ell \leq G, \quad 1 \leq r \leq C, \quad 1 \leq k \leq E_r. \tag{2.15}$$

2. *Rate de-matching.*
The rate de-matching aims to covert the code block message to the format that can be used for 5G LDPC parity-check matrix to process decoding. Rate de-matching is done on each code block independently. Assume that the input message is $f_r(0), f_r(1), \ldots, f_r(E_r - 1)$. The output message from rate de-matching is $g(0), g(1), \ldots, g(N_r + 2\mathrm{Z_c} - 1)$

$$g_k = \begin{cases} 0, & \texttt{for} \ 1 \leq k \leq 2\mathrm{Z_c}, \\ f_k, & \texttt{for} \ 2\mathrm{Z_c} + 1 \leq k \leq E_r, \\ 0, & \texttt{for} \ E_r + 1 \leq k \leq \mathrm{N} + 2\mathrm{Z_c}. \end{cases} \tag{2.16}$$

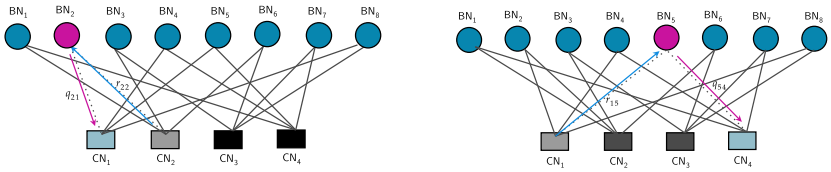
3. *LDPC decoding*.
The LDPC decoding is done on each code block independently, and many decoding algorithms can be used. Subsection 1.2.4 highlights different LDPC decoding algorithms.

4. Code block de-segmentation. The CB de-segmentation is used to extract the message bits and transport block attached CRC bits. Assume that the output from code block de-segmentation is $\hat{b}(0), \hat{b}(1), \ldots, \hat{b}(B-1)$, where B is the size of original transport block information with attached CRC bits. The input to the code block de-segmentation is $h_r(0), h_r(1), \ldots, h_r(\mathrm{N} + 2\mathrm{Z_c} - 1)$



(a) $\mathsf{CN}_1 \longrightarrow \mathsf{BN}_2$. (b) $\mathsf{CN}_4 \longrightarrow \mathsf{BN}_5$.

Figure 2.4: Check Node Updates



(a) $\mathsf{BN}_2 \longrightarrow \mathsf{CN}_1$. (b) $\mathsf{BN}_5 \longrightarrow \mathsf{CN}_4$.

Figure 2.5: Bit Node Updates

$$\hat{b}_k = h_{r,s}, \quad \texttt{where } 1 \leq k \leq B, \quad 1 \leq r \leq C, \quad 1 \leq s \leq \mathrm{K}_r - L. \tag{2.17}$$

Where K_r is the number of information and CRC bits in a code block and L is the length of CRC bits in a code block.

5. *CRC check*:
   The CRC detachment is used to extract the CRC bits in transport block after the information transmitted in 5G NR shared channels. Then the extracted CRC bits will be checked with the original CRC bits attached to transport block information before transmitted.

1.2.4 Soft-Decision based LDPC decoding Algorithms

The optimal performing method is soft-decision decoding, involving the computation of *log-likelihood ratios* (LLRs) and the exchange of extrinsic information between variable and parity nodes. This method is known by various names in the literature, such as the *belief propagation algorithm* (BPA), or *message massing algorithm* (MPA), or *sum product* (SPA) algorithm. Additionally, there is the *min-sum algorithm*(MSA), an approximate method with lower complexity compared to SPA. Various algorithmic variants are available, tailored to specific practical applications and necessitating simplifications for tractable implementation.

The message passing decoding can be divided into bit or variable nodes' operation, also called row operation, and check nodes' operation, also called column operation. A MPA based on Pearl's belief algorithm describes the iterative decoding steps. The message probability passed between check nodes and variable nodes can be called belief, such as q_{ij} and r_{ji} in Figures 2.4 and 2.5.

Index sets $\mathsf{BN}_{j\setminus i}$ and $\mathsf{CN}_{i\setminus j}$ are based on the *parity check matrix* (PCM). Index set CN_i and BN_j correspond to all non-zero element on column i and row j of the PCM, respectively. Figure 2.6 a simple conceptual illustration of BN and CN index sets within the PCM provided in (2.1) for the specified values of $i = 3$ and $j = 2$. Furthermore, the beliefs can be described via the following equations. Assume that a sequence of information bits are independently $\mathbf{c} = [\mathrm{c}(0), \mathrm{c}(1), \ldots, \mathrm{c}(\mathrm{N}-1)]$, and consider the following notation :

- BN_j = Bit nodes connected to check node j,
- $\mathsf{BN}_{j\setminus i}$ =bit nodes connected to check node j, excluding bit node i,
- CN_i = check nodes connected to variable node i,
- $\mathsf{CN}_{i\setminus j}$ =check nodes connected to bit node i, excluding check node j,
- $p_i = p_r(\mathrm{c}_i = 1|\mathrm{y}_i)$, is the probability of $\mathrm{c}_k = 1$.
- y_i is the channel sample at variable node i.
- The check-to-variable extrinsic message passing to the $j-th$ variable node from the $i-th$ check node is denoted by $\mathscr{L}(\mathrm{r}_{ji})$ and $\mathscr{L}(\mathrm{q}_{ij})$ is the variable-to-check extrinsic message.

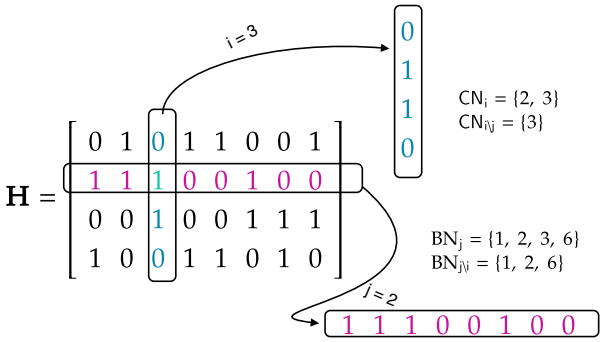


Figure 2.6: Conceptual Illustration of CNs and BNs processing Within BN and CN index sets for the specified values of i and j.

The belief propagation algorithm is adaptable to representation in both probability and log domains, wherein probabilities are expressed as LLRs. Employing LLR domain decoding offers a reduction in implementation complexity, since multiplications in the probability domain can be equivalently represented as additions in the log domain. Besides, many multiplications of probabilities involved could become numerically unstable, so the log domain algorithm is preferred [85].

1. **Sum Product Algorithm (SPA)**:
   The Sum Product Algorithm, also known as (BPA), constitutes a fundamental soft decision decoding approach where messages are conveyed as probabilities. The implementation of Belief Propagation relies on the decoding algorithm introduced by Gallager [10]. For a transmitted LDPC encoded codeword, $\mathbf{c}$, the input to the LDPC decoder is the LLR value defined as follows :

$$\mathscr{L}\left(\mathrm{c}_i\right)=\log \frac{p_r\left(\mathrm{c}_i=0 \mid \texttt{channel output for}\ \mathrm{c}_i\right)}{p_r\left(\mathrm{c}_i=1 \mid \texttt{channel output for}\ \mathrm{c}_i\right)}. \tag{2.18}$$

   In each iteration, the algorithm updates its key components through horizontal and vertical processing steps. The check nodes to bit nodes operation (*horizontal processing*) is based on (2.19).

$$\begin{aligned}\mathscr{L}\left(\mathrm{r}_{ji}\right)=\log \frac{\mathrm{r}_{ji}(0)}{\mathrm{r}_{ji}(1)} &= 2 \tanh^{-1}\left(\prod_{i' \in \mathsf{BN}_{j \setminus i}} \tanh\left(\frac{1}{2} \mathscr{L}\left(\mathrm{q}_{i'j}\right)\right)\right), \\ &= \left(\prod_{i' \in \mathsf{BN}_{j \setminus i}} \operatorname{sign}\left(\mathscr{L}\left(\mathrm{q}_{i'j}\right)\right)\right) \phi\left(\sum_{i' \in \mathsf{BN}_{j \setminus i}} \phi\left(\left|\mathscr{L}\left(\mathrm{q}_{i'j}\right)\right|\right)\right).\end{aligned} \tag{2.19}$$

   Where $\phi(\mathbf{x})=-\log\left[\tanh\left(\frac{|\mathbf{x}|}{2}\right)\right]=\log\left(\frac{\mathrm{e}^{\mathbf{x}}+1}{\mathrm{e}^{\mathbf{x}}-1}\right)$.
   The bit nodes to check nodes' operation (*vertical processing*)is given by.

$$\mathscr{L}\left(\mathrm{q}_{ij}\right)=\mathscr{L}\left(\mathrm{c}_i\right)+\sum_{j' \in \mathsf{CN}_{i \setminus j}} \mathscr{L}\left(\mathrm{r}_{j'i}\right). \tag{2.20}$$

$$\mathscr{L}\left(Q_i\right)=\mathscr{L}\left(\mathrm{c}_i\right)+\sum_{j' \in \mathsf{CN}_i} \mathscr{L}\left(\mathrm{r}_{j'i}\right). \tag{2.21}$$

   Where $L(Q_i)$ is the output LLR from the decoder and can be used to make decision.

$$\hat{\mathrm{c}}_i=\begin{cases}1, & \text{if } \mathscr{L}\left(Q_i\right)<0, \\ 0, & \text{else .}\end{cases} \tag{2.22}$$

   Repeat the steps until the maximum iterations are done or $\mathbf{H}\hat{\mathbf{c}}^{\mathsf{T}}=\mathbf{0}$.

Algorithm 1: Log-Likelihood Belief Propagation Decoding Algorithm

Input: The channel log likelihoods : $\mathscr{L}_i \in \mathbb{R}^n$:
Maximum # of iterations, MAXITER
Description of the parity check matrix using $\mathsf{BN}(j)$ and $\mathsf{CN}(i)$.
Ouput: Estimated code word: $\hat{\mathbf{c}} \in \{0,1\}^n$.
`Initialization :`
for *each i, and for each j* $\in \mathsf{CN}(i)$ **do**
  $\mathscr{L}\left(\mathrm{q}_{ij}\right) = \mathscr{L}_i$
`Check Node to Variable Node Step (horizontal step):`
for *each check node* j **do**
  for *each variable node* $i \in \mathsf{BN}(j)$ **do**
    $$\mathscr{L}\left(\mathrm{r}_{ji}\right) = 2\tanh^{-1}\left(\prod_{i' \in \mathsf{BN}_{j\setminus i}} \tanh\left(\frac{1}{2}\mathscr{L}\left(\mathrm{q}_{i'j}\right)\right)\right)$$
`Variable Node to Check Node Step (vertical step)`
for *each variable node* i **do**
  for *each check node* $j \in \mathsf{CN}(i)$ **do**
    $$\mathscr{L}\left(\mathrm{q}_{ij}\right) = \mathscr{L}_i + \sum_{j' \in \mathsf{CN}_{i\setminus j}} \mathscr{L}\left(\mathrm{r}_{j'i}\right)$$
    Also compute the output likelihoods
    $$\mathscr{L}\left(\mathrm{Q}_i\right) = \mathscr{L}_i + \sum_{j' \in \mathsf{CN}_i} \mathscr{L}\left(\mathrm{r}_{j'i}\right)$$
`Hard decision:`
for *each* i **do**
  $\hat{\mathrm{c}}_i = 1$ if $\mathscr{L}\left(\mathrm{Q}_i\right) < 0$ else $\hat{\mathrm{c}}_i = 0$
`Parity Check:`
if $\mathbf{H}\hat{\mathbf{c}}^{\mathsf{T}} = \mathbf{0}$ **then**
  return $\hat{\mathbf{c}}$
otherwise, **if** # ITER $<$ MAXITER
goto Check Node to Variable Node Step
else
  return $\hat{\mathbf{c}}$ and indication of coding failure.

The BP algorithm achieves near-optimal decoding performance, but suffers from high computational complexity. In order to find a better trade-off between performance and complexity, a number of efficient decoding algorithms have been proposed in the scientific literature.

2. **Min-Sum Algorithm (MSA)**:
Min-Sum Algorithm (MSA) for LDPC decoding is a reduced complexity decoding algorithm with min-sum approximation compared to sum product algorithm or belief propagation algorithm. Indeed, the value of $\phi(x)$ decreases sharply to almost when x increases. So the smallest $|\mathscr{L}\left(\mathrm{q}_{i'j}\right)|$ value dominates the summation $\sum_{i' \in \mathsf{BN}_{j\setminus i}} \phi\left(|\mathscr{L}\left(\mathrm{q}_{i'j}\right)|\right)$. Thus, it comes

$$\sum_{i' \in \mathsf{BN}_{j\setminus i}} \phi\left(|\mathscr{L}\left(\mathrm{q}_{i'j}\right)|\right) \approx \phi\left(\min_{i\prime \in \mathsf{BN}_{j\setminus i}} \left(|\mathscr{L}\left(\mathrm{q}_{i'j}\right)|\right)\right). \tag{2.23}$$

A such approximation is used for MSA. For MSA, other computations are the same with SPA, except $\mathscr{L}\left(\mathrm{r}_{ji}\right)$.

$$\mathscr{L}\left(\mathrm{r}_{ji}\right) = \left(\prod_{i\prime \in \mathsf{BN}_{j\setminus i}} \operatorname{sign}\left(\mathscr{L}\left(\mathrm{q}_{i'j}\right)\right)\right) \cdot \min_{i\prime \in \mathsf{BN}_{j\setminus i}} \left(|\mathscr{L}\left(\mathrm{q}_{i'j}\right)|\right). \tag{2.24}$$

3. **Optimized Min-Sum Decoding Algorithm**:

It is shown that the magnitude of $\phi\left(\min_{i'}\left(\left|\mathscr{L}\left(\mathrm{q}_{i'j}\right)\right|\right)\right)$ obtained in MSA is always greater than the magnitude of $\sum_{i'\in \mathsf{BN}_{j\setminus i}} \phi\left(\left|\mathscr{L}\left(\mathrm{q}_{i'j}\right)\right|\right)$ obtained in SPA. Outputs from variable nodes in MSA decoding are overestimated compared to SPA due to the approximation. There are several methods to optimize MSA to make the approximation more accurate. The two most popular methods are *normalized min-sum algorithm(NMSA)* and *offset min-sum algorithm(OMSA)* and are presented in [46]. The idea behind NMSA and OMSA is to reduce the magnitude of variable node outputs.

(a) Normalized Min-Sum Algorithm, (2.24) simply becomes:

$$\mathscr{L}_{\text{NMSA}}\left(\mathrm{r}_{ji}\right)=\left(\prod_{i'\in \mathsf{BN}_{j\setminus i}} \operatorname{sign}\left(\mathscr{L}\left(\mathrm{q}_{i'j}\right)\right)\right)\cdot \min_{i'\in \mathsf{BN}_{j\setminus i}}\left(\alpha\cdot\left|\mathscr{L}\left(\mathrm{q}_{i'j}\right)\right|\right), \tag{2.25}$$

where α is called normalization or scaling factor, $\alpha\in]0,1)$.

(b) Offset Min-Sum Algorithm, (2.24) becomes:

$$\mathscr{L}_{\text{OMSA}}\left(\mathrm{r}_{ji}\right)=\left(\prod_{i'\in \mathsf{BN}_{j\setminus i}} \operatorname{sign}\left(\mathscr{L}\left(\mathrm{q}_{i'j}\right)\right)\right)\cdot \max\left(\min_{i'\in \mathsf{BN}_{j\setminus i}}\left(\left|\mathscr{L}\left(\mathrm{q}_{i'j}\right)\right|-\beta\right),0\right), \tag{2.26}$$

where $\beta\geq 0$ is the offset value.

4. **Layered Belief Propagation Algorithm**:
Layered belief propagation (LBP) algorithm is an adaptation of the decoding algorithm presented in [47]. In the LBP algorithm, the decoder executes CNPs based on a node-by-node mode until all check functions are satisfied, or the iteration reaches the maximum value [48]. The decoding loop iterates over subsets of rows (*layers*) of the PCM.

A CNP is composed of a series of operations that update the values of $\mathscr{L}\left(\mathrm{Q}_i\right)$ and $\mathscr{L}\left(\mathrm{r}_{ji}\right)$ as follows:

$$\begin{aligned}
&\texttt{(1) Update Input LLRs} && \mathscr{L}\left(\mathrm{q}_{ij}\right)=\mathscr{L}\left(\mathrm{Q}_i\right)-\mathscr{L}\left(\mathrm{r}_{ji}\right).\\
&\texttt{(2) Perform CNP} && \mathscr{L}'\left(\mathrm{r}_{ji}\right)=2\tanh^{-1}\left(\prod_{i'\in \mathsf{BN}_{j\setminus i}}\tanh\left(\frac{1}{2}\mathscr{L}\left(\mathrm{q}_{i'j}\right)\right)\right).\\
&\texttt{(3) Perform Output LLRs} && \mathscr{L}'\left(\mathrm{Q}_i\right)=\mathscr{L}\left(\mathrm{q}_{ij}\right)+\mathscr{L}'\left(\mathrm{r}_{ji}\right).
\end{aligned} \tag{2.27}$$

For each layer, the decoding stage (3) works on the combined input obtained from the current LLR inputs $=\mathscr{L}\left(\mathrm{q}_{ij}\right)$ and the previous layer updates $\mathscr{L}'\left(\mathrm{r}_{ji}\right)$.

Because only a subset of the nodes is updated in a layer, the layered belief propagation algorithm is faster compared to the belief propagation algorithm. As shown in [48], the convergence speed of the LBP algorithm is about twice as fast as that of the BP algorithm. To achieve the same error rate as attained with belief propagation decoding, use half the number of decoding iterations when using the layered belief propagation algorithm.

In the layered decoding approach, each layer operates on variable nodes and check nodes independently. The input *log-likelihood ratio* (LLR) for a given layer is derived from the output LLR of the preceding layer.

Ultimately, the output LLR of the final layer serves as the output LLR of the decoding process, thereby informing the decision-making process.

1.3 5G NR Polar Codes

Polar codes were first proposed by Arikan [50] in 2009. It has quickly become a research hot spot in the coding community, with the advantages of theoretical accessibility to the Shannon limit and simple compiled code algorithms. Using polar codes as the channel coding scheme for 5G control channels [132, 133] has demonstrated the significance of Arikan's invention, and its applicability in commercial systems has been proven. This coding family achieves capacity rather than merely approaching it, since it is based on the idea of channel polarization. Moreover, polar codes can be used for any code rate and for any code lengths shorter than the maximum code length due to their adaptability.
Polar codes are the first type of forward error correction codes achieving the symmetric capacity for arbitrary binary-input discrete memoryless channel under low-complexity encoding and low-complexity *successive cancellation* (SC)

decoding with order of $\mathcal{O}(N \log N)$ for infinite length codes. Polar codes are founded based on several concepts including channel polarization, code construction, polar encoding, which is a special case of the normal encoding process (i.e., more structural) and its decoding concept [51].

The polar code is a type of block code, *but it is not a linear code*. Polar codes are constructed from non-linear transformations called polar transformations. They exploit properties of certain transformations to make certain parts of the code carry useful information, while other parts act as frozen bits whose value is fixed. They are constructed using the polar channel transform, generally based on the Hadamard transform, and have the particularity of approaching the capacity limit of the communication channel efficiently.

In short, although polar codes are block codes, they differ from more conventional linear codes and use non-linear transformations to achieve decoding performance close to the theoretical channel limit. Furthermore, from the standpoint of 5G's physical channels, control information is typically transmitted with a relatively small number of information bits and a small block width, so a low coding rate with good performance in a lower BLER is required, and polar codes can meet this requirement.

1.3.1 State-of-art Polar codes

Polar codes have emerged as a key channel coding scheme for 5G NR [60]. The trend in polar code design and decoding techniques has been driven by the need for efficient and reliable communication. These codes exhibit promising characteristics such as rate flexibility and low decoding latency, addressing crucial requirements for 5G systems. However, challenges persist, particularly in reducing latency without compromising reliability [61]. Efforts have been made to enhance the decoding process, such as the development of low latency decoders for short block length polar codes. Thus, Gamage *et al.* [61] highlighted decoders that utilize simplified SC algorithms combined with list decoding techniques to achieve the desired balance between reliability and latency in ultra-reliable low-latency communication systems. Subsequently, Geiselhart *et al.* [62] leveraged CRC codes to aid belief propagation list decoding, improving error-rate performance while optimizing decoding complexity. Meanwhile, Piao *et al.* [63] proposed innovative decoding algorithms like CRC-aided sphere decoding to enhance the performance of short polar codes. By utilizing CRC information, these algorithms provide stable performance across various code rates. Additionally, Cavatassi *et al.* [65] introduced asymmetric coding schemes to allow for arbitrary block lengths, reducing decoding complexity while maintaining error correction performance. Shen *et al.* [64] explored fast iterative soft-output list decoding to improve error-rate performance and decoding efficiency. Kaykac *et al.* [66] highlighted that understanding the operation and performance of 5G polar codes is crucial as they are integral to the functionality of 5G control channels. Ercan *et al.* [67] optimized practical implementations of polar code decoders, developing dynamic SC-flip decoding algorithms with reduced complexity. Kestel *et al.* [68] quantified trade-offs between error-correction capability and implementation costs, crucial for achieving efficient high-throughput decoding in 5G systems. Moreover, Arli and Gazi [69] and Sun *et al.* [71] proposed adaptive belief propagation algorithms and low-complexity decoding schemes, respectively, to address challenges in decoding polar codes efficiently for high throughput applications.

Moreover, the advancements in polar code decoding algorithms and software and hardware architectures have paved the way for efficient and low-latency implementations, including 5G and beyond. Sarkis *et al.* [76] introduced a framework for generating high-speed software polar decoders, achieving significant throughput improvements. Meanwhile, in [77], simplified decoding algorithms have been proposed to enhance the speed of polar list decoders, maintaining error-correction performance. subsequently, Kam *et al.* [78] addressed the latency issue in SC decoding by introducing tree-level parallelism and novel pruning methods, significantly reducing decoding latency. Moreover, Xiang *et al.* [73] presented a reduced-complexity logarithmic SC stack (Log-SCS) polar decoding algorithm, achieving notable improvements in decoding latency and complexity. Additionally, Rezaei *et al.* [75] focused on implementing ultra-fast polar decoders with new sub-codes and decoding algorithms for short to moderate block lengths, emphasizing hardware optimization techniques. Furthermore, Liu *et al.* [74] proposed high-throughput adaptive list decoding architectures for polar codes on GPUs, leveraging adaptive mapping strategies to improve throughput and latency performance.

Recently, research in decoding algorithms, combining techniques from deep learning, reinforcement learning, and traditional decoding methods, offer promising solutions for efficient and low-latency decoding of polar codes and is among the hot topics in the area. Doan *et al.* [79] introduced a novel approach, *neural belief propagation* (NBP), combining CRC with polar codes to enhance error-correction performance, particularly for parallel iterative BP decoders. Building upon this, Hashemi *et al.* [80] proposed a deep-learning-aided successive-cancellation list (DL-SCL) decoding algorithm, leveraging deep learning techniques to optimize bit-flipping metrics and reduce computational complexity.

Meanwhile, Doan *et al.* [81] addressed factor-graph permutation selection in polar codes using reinforcement learning, achieving significant error-correction performance gains and expanding on this framework, fast SC-flip decoding, a bit-flipping algorithm optimized via reinforcement learning have been introduced in Doan *et al.* [82] for improved error-correction performance in polar codes. Additionally, Wodiany and Pop [83] presented a low-precision *neural network* (NN) decoder to mitigate the scalability issues and high memory usage of conventional NN decoders, maintaining wireless performance with reduced computational complexity. In parallel, Wen *et al.* [84] proposed a BP-NN decoding algorithm for polar codes, integrating neural network decoders into the belief propagation framework to reduce decoding delay while maintaining low bit error rates. These advancements in decoding algorithms, combining techniques from deep learning, reinforcement learning, and traditional decoding methods, offer promising solutions for efficient and low-latency decoding of polar codes in future communication systems.
Furthermore, the primary polar code decoding algorithms include the SC algorithm [50], the SCL algorithm [52, 53], the CA-SCL algorithm [54], the BP algorithm [56], and the SCAN algorithm [57]. Originally proposed by Arikan, the SC algorithm's performance diminishes for finite length codes. SCL, an enhancement of SC, offers superior performance by providing multiple paths. CA-SCL, incorporating cyclic redundancy checks on message bits over SCL, significantly boosts performance through simple checksums. Currently, 3GPP polar decoding relies on the CA-SCL algorithm, surpassing LDPC codes. Notably, SC, SCL, and CA-SCL algorithms are hard output, yielding bit sequences rather than LLR values. To facilitate joint designs, soft output algorithms providing LLR values are essential; BP and SCAN algorithms serve this purpose. Decoding delay varies between BP and SCAN: BP utilizes the "*flood*" rule for message passing [59], while SCAN employs the serial elimination rule for SC-like algorithms. BP exhibits lower decoding delay, whereas SCAN demonstrates superior convergence speed. In terms of performance, these algorithms rank as follows: $\texttt{CA-SCL} > \texttt{state of the art (LDPC, Turbo)} > \texttt{SCL} > \texttt{BP} = \texttt{SCAN} > \texttt{SC}$.

1.3.2 Foundations and Fundamentals

A polar code of length $\mathrm{N} = 2^n$ is generated using a generator matrix $\mathbf{G}$ of size $\mathrm{N} \times \mathrm{N}$. A block of length N, consisting of $\mathrm{N} - \mathrm{K}$ frozen bits and K information bits, is multiplied by $\mathbf{G}$ to produce the polar codeword $\mathbf{x} = \mathbf{uG}$. The generation matrix can be expressed as follows:

$$\mathbf{G}_{\mathrm{N}} = \mathbf{B}_{\mathrm{N}}\mathbf{F}^{\otimes n}, \tag{2.28}$$

where $\mathbf{B}_{\mathrm{N}}$ is the bit-reversal permutation matrix, $\mathbf{F}^{\otimes n}$ is the $n - fold$ Kronecker product of $\mathbf{F}$ with itself, defined recursively as

$$\mathbf{F}^{\otimes 1} = \mathbf{F} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{F}^{\otimes n} = \begin{bmatrix} \mathbf{F}^{\otimes n-1} & \mathbf{0} \\ \mathbf{F}^{\otimes n-1} & \mathbf{F}^{\otimes n-1} \end{bmatrix}. \tag{2.29}$$

The encoding operation can be expressed as

$$\mathrm{x}_0^{\mathrm{N}-1} = \left(\mathrm{u}_0^{\mathrm{N}-1}\right)\mathbf{B}_{\mathrm{N}}\mathbf{F}^{\otimes n}, \tag{2.30}$$

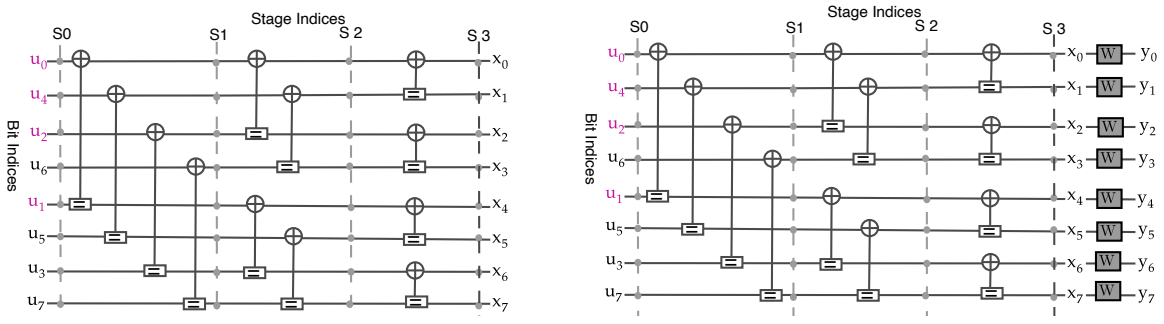
where $\otimes$ denotes the Kronecker product.

Remark 2.1. The presence of the *bit-reverse permutation matrix* $\mathbf{B}_{\mathrm{N}}$ doesn't impact the code's distance properties. Some implementations omit it. When the bit-reverse permutation isn't utilized, the encoder is referred to as being in *natural order* [85].

For instance, considering a polar code of length $\mathrm{N} = 8$, once the positions of the fixed bits, termed *frozen bits*(i.e., magenta color), and the positions of the information bits(i.e., dark color) are determined, the encoder graph is depicted in Figure 2.7a. Consequently, the input/output relationship in (2.31) is established according to the encoder graph representation shown in Figure 2.7a.

$$\begin{aligned} \mathrm{x}_0 &= \mathrm{u}_0 \oplus \mathrm{u}_1 \oplus \mathrm{u}_2 \oplus \mathrm{u}_4, & \mathrm{x}_1 &= \mathrm{u}_4 \oplus \mathrm{u}_5 \oplus \mathrm{u}_6, \\ \mathrm{x}_4 &= \mathrm{u}_1 \oplus \mathrm{u}_3 \oplus \mathrm{u}_6, & \mathrm{x}_5 &= \mathrm{u}_5 \oplus \mathrm{u}_7, \\ \mathrm{x}_2 &= \mathrm{u}_2 \oplus \mathrm{u}_3 \oplus \mathrm{u}_6, & \mathrm{x}_3 &= \mathrm{u}_6 \oplus \mathrm{u}_7, \\ \mathrm{x}_6 &= \mathrm{u}_3 \oplus \mathrm{u}_7, & \mathrm{x}_7 &= \mathrm{u}_7, \end{aligned} \tag{2.31}$$

where $\oplus$ denotes addition in $\mathbb{F}_2$. The codeword $[\mathrm{x}_0, \mathrm{x}_1, \ldots, \mathrm{x}_7]$ is transmitted through the physical channels W^8, with channel outputs $[\mathrm{y}_0, \mathrm{y}_1, \ldots, \mathrm{y}_7]$. This structure, comprising coding and channel transmission, creates the channel $\mathrm{W}^8 : \mathrm{x}_0, \mathrm{x}_1, \ldots, \mathrm{x}_7 \longmapsto \mathrm{y}_0, \mathrm{y}_1, \ldots, \mathrm{y}_7$.



(a) Polar encoder graph of size N=8.

(b) Polar encoder of size N=8 over hysical channel W.

Figure 2.7: The graph of an $(8, 4)$ polar code with bit reversed indexing transmitted over physical channels (W).

Initially, polar codes were nonsystematic, but they can be transformed into systematic codes like any linear code. Systematic polar encoding utilizes the standard non-systematic polar encoding apparatus. Systematic polar codes offer improved BER performance compared to non-systematic ones, yet both have identical BLER performance [18]. Systematic polar coding demonstrates greater resilience to error propagation with SC decoder than non-systematic polar coding.

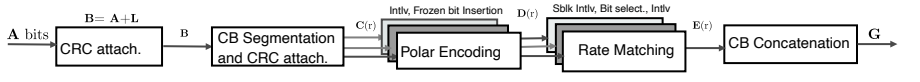
In a linear code, a codeword is a point in the row space of the generator matrix $\mathbf{G}$, so that in $\mathbf{x} = \mathbf{z}\mathbf{G}$, $\mathbf{z}$ is a codeword, regardless of the particular values in $\mathbf{z}$. One way to do encoding might be to take a message vector $\mathbf{u}$, place it into K elements of the codeword $\mathbf{x}$, then find a vector $\mathbf{z}$ which fill in the remaining $\mathrm{N} - \mathrm{K}$ elements of $\mathbf{x}$ in such a way that $\mathbf{x}$ is in the row space of $\mathbf{G}$. If $\mathbf{z}$ can be found via linear operations from $\mathbf{u}$ in such a way that the message symbols appear explicitly in $\mathbf{u}$, then systematic encoding has been achieved [85].

$$\mathbf{G}_8 = \mathbf{F}^{\otimes 3} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{2.32}$$

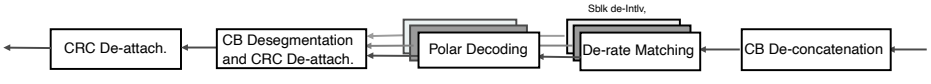
Note that we can perform systematic encoding via the *encoder Graph, Arikan's method, the bit reverse permutation, etc.* [85].

1.3.3 3GPP 5G Polar Codes

In NR, the polar code is used to encode *broadcast channel* (BCH) as well as *downlink control information* DCI and *uplink control information* (UCI). The overall control streams transmission chain from MAC/PHY layer processing schematic is depicted in Figure 2.8, which describes the transmit-end for the physical uplink/downlink control channel supporting a transport block CRC attachment, code block segmentation and code block CRC attachment, Polar encoding, rate matching code block concatenation. The receiving chain is in line with the transmitting chain in the reverse flow.



(a) Tranmitter end.



(b) Receiver end.

Figure 2.8: Conceptual illustration of 5G Polar transceiver chain.

3GPP NR uses a variant of the polar code called *distributed* CRC (D-CRC) polar code, that is, a combination of *CRC-assisted and* polar codes *(PC*, which interleaves a CRC-concatenated block and relocates some of the PC bits into

the middle positions of this block prior to performing the conventional polar encoding[18]. This allows a decoder to early terminate the decoding process as soon as any parity check is not successful. *The D-CRC scheme is important for early termination of decoding process*, because the post-CRC interleaver can distribute information and CRC bits such that partial CRC checks can be performed during list decoding and paths failing partial CRC check can be pruned, leading to early termination of decoding. The post-CRC interleaver design is closely tied to the CRC generator polynomial, thus by appropriately selecting the CRC polynomial, one can achieve better early termination gains and maintain acceptable false alarm rate.

1. *CRC attachment*.
   Assume that the input message (control information) before CRC attachment is $a(0), a(1), \ldots, a(A-1)$, where A is input sequence, parity bits are $p(0), p(2), \ldots, p(L-1)$, L is the number of parity bits. The parity bits are generated by one of the following cyclic generator polynomials.
   A CRC length $L = 24$ bits is utilized for the downlink, and depending on the amount of A, CRCs of $L = 6$ and $L = 11$ bits are provided for the uplink.
   For downlink channels, the generator polynomial $\mathrm{g}_{\mathsf{CRC24A}}(\beta)$ is used.

$$\begin{aligned}\mathrm{g}_{\mathsf{CRC24A}}(\beta) = \big[&\beta^{24} + \beta^{23} + \beta^{18} + \beta^{17} + \beta^{14} + \beta^{11} + \beta^{10} \\ &+\beta^{7} + \beta^{6} + \beta^{5} + \beta^{4} + \beta^{3} + \beta + 1\big].\end{aligned} \tag{2.33}$$

   And for uplink channels, the generator polynomial $\mathrm{g}_{\mathsf{CRC11}}(\beta)$ or $\mathrm{g}_{\mathsf{CRC6}}(\beta)$ is used.

$$\mathrm{g}_{\mathsf{CRC11}}(\beta) = \big[\beta^{11} + \beta^{10} + \beta^{9} + \beta^{5} + 1\big], \quad \mathrm{g}_{\mathsf{CRC6}}(\beta) = \big[\beta^{6} + \beta^{5} + 1\big]. \tag{2.34}$$

   The message bits after attaching CRC are $b(1), b(2), \ldots, b(B)$, B is the size of transport block information with CRC bits and $B = A + L$.

$$b_k = \begin{cases} a_k, & \texttt{for } k = 0, 1, \ldots, A-1 \\ p_{k-A}, & \texttt{for } k = A, A+1, \ldots, A+L-1 \end{cases} \tag{2.35}$$

2. *Code block segmentation and code block CRC attachment*.
   The input bit sequence to the code block segmentation is denoted $a(0), a(1), \ldots, a(A-1)$, where A is no larger than 1706. Assume that the maximum code block size is A', assume that C is the total number of code blocks. Thus,

$$A' = \lceil A/C \rceil \cdot C. \tag{2.36}$$

   The sequence $c_r(0), c_r(1), \ldots, c_r(A'/C-1)$ is used to calculate the CRC parity bits $p_r(0), p_r(1), \ldots, p_r(L-1)$, such that

$$c_{r,k} = p_{r,(k-A'/C)}, \quad A'/C \le k \le A'/C + L - 1, \quad 1 \le r \le C \tag{2.37}$$

   At the transmitter end, we have the following streamlines:
   The bit sequence input for a given code block to channel coding is denoted by $c(0), c(1), \ldots, c(\mathrm{K}_r - 1)$, where K_r is the number of bits in code block number r, and each code block is individually encoded. After the encoding process, the resulting coded bit sequence within the $r-th$ code block is denoted by $d_r(0), d_r(1), \ldots, d_r(\mathrm{N}_r - 1)$ where $\mathrm{N}_r = 2^n$ (code length of the polar code) determined by the following:
   `if` $E_r \le (9/8) \cdot 2^{(\lceil \log_2 E_r \rceil - 1)}$ `and` $\mathrm{K}_r / E_r < 9/16$,
   $n_1 = \lceil \log_2 E_r \rceil - 1$.
   `else` $n_1 = \lceil \log_2 E_r \rceil$.
   $\mathrm{r}_{\min} = 1/8$; $n_2 = \lceil \log_2 (K/\mathrm{r}_{\min}) \rceil$;
   $n = \max\{\min\{n_1, n_2, n_{\max}\}, n_{\min}\}$,
   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. E_r is the rate matching output sequence length.
   UE is not expected to be configured with $\mathrm{K}_r + n_{PC} > E$, where n_{PC} is the number of parity check bits.

   - *Interleaving*.
     The bit sequence $c_r(0), c_r(1), \ldots, c_r(\mathrm{K}_r - 1)$ is interleaved into bit sequence $c'_r(0), c'_r(1), \ldots, c'_r(\mathrm{K}_r - 1)$

a follows:

$$c'_{r,k} = c_{r,\Pi(k)}, k = 0, 1, \ldots, \mathrm{K}_r - 1, \quad r = 0, 1, \ldots, C-1 \tag{2.38}$$

where $\Pi(k)$ is the interleaving pattern [132].

- *Polar encoding*.
  The interleaved vector $\mathbf{c}'$ is assigned to the information set along with the PC bits, while the remaining bits in the N-bit vector $\mathbf{u}$ are frozen. Hence, $\mathbf{u} = u(0), u(1), \ldots, u(\mathrm{N}-1)$ is generated according to the clause 5.3.1.2 [132]. Denote $\mathbf{G}_{\mathsf{N}_r} = (\mathbf{G}_2)^{\otimes n}$ 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 $\mathbf{d}_r = d_r(0), d_r(1), \ldots, d_r(\mathrm{N}_r - 1)$ is obtained by $\mathbf{d}_r = \mathbf{u}_r \mathbf{G}_{\mathsf{N}_r}$, where encoding is performed in $\mathbb{F}_2$.

3. *Rate matching*.
   The rate matching for polar code is defined per coded block and consists of *sub-block interleaving*, bit collection, and bit interleaving. Sequence of coded bits at the rate matcher input is $d_r(0), d_r(1), \ldots, d_r(\mathrm{N}_r - 1)$, The output bit sequence from the rate matcher is denoted as $f_r(0), f_r(1), \ldots, f_r(E-1)$. For rate matching, puncturing, shortening ($E_r < \mathrm{N}_R$), or repetition($E_r \geq N_r$) are applied to change the N_r-bit vector $\boldsymbol{s}_r$ into the E_r-bit vector $\boldsymbol{e}_r$.
   Indeed, the rate matching process encapsulates the following steps:

   - *Sub-block interleaving*.
     The bits input to the sub-block interleaver are the coded bits $d_r(0), d_r(1), \ldots, d_r(\mathrm{N}_r - 1)$. The coded bits $d_r(0), d_r(1), \ldots, d(\mathrm{N}_r - 1)$ are divided into 32 sub-blocks. The bits output from the sub-block interleaver are denoted as $s_r(0), s_r(1), \ldots, s_r(\mathrm{N}_r - 1)$.
   - *Bit selection*.
     The bit sequence after the sub-block interleaver $s_r(0), s_r(1), \ldots, s_r(\mathrm{N}_r - 1)$ is written into a circular buffer of length N. Denoting by E_r the rate matching output sequence length, the bit selection, output bit sequence $e_r(0), e_r(1), \ldots, e_r(\mathrm{E}_r - 1)$.
   - *Interleaving of coded bits*.
     The bits sequence $e(0), e(1), \ldots, e(\mathrm{E}_r - 1)$ is interleaved into bit sequence $f_r(0), f_r(1), \ldots, f_r(E-1)$., where the value of E_r is no larger than 8192.

4. *Code block concatenation*.
   The code block concatenation consists of sequentially concatenating the rate matching outputs for the different code blocks.
   The input bit sequence for the code block concatenation block are the sequences $f_{r,k}$, for $r = 0, \ldots, C-1$ and $\mathrm{k} = 0, \ldots, E_r - 1$, where E_r is the number of rate matched bits for the $r-th$ code block. The output bit sequence from the code block concatenation block is the sequence g_ℓ for $\ell = 0, \ldots, G-1$. Therefore,

$$g_\ell = f_{r,k}, \texttt{where} \quad 1 \leq \ell \leq G, \quad 1 \leq \mathrm{r} \leq C, 1 \leq k \leq E_r. \tag{2.39}$$

At the receiver end, the procedure is as follows:

1. *Code block de-concatenation*.
   Assume that the input message to code block de-concatenation is $y(0), y(1), \ldots, y(G-1)$. The output message from code block de-concatenation is $f_{r,k}$, where $\quad 1 \leq \ell \leq G, \quad 1 \leq \mathrm{r} \leq C, 1 \leq k \leq E_r$.
2. *Rate de-matching*.
   The purpose of rate de-matching is to convert the code block message to the format that can be used for 5G Polar decoder to process decoding. Rate de-matching is done on each code block independently.
3. *Polar decoding*.
   The decoding process is done on each code block independently. The subsequent subsection 1.3.4 provides more details on polar decoding algorithms.
4. *Code block de-segmentation*.
   Assume that the output from code block de-segmentation is $\hat{b}(0), \hat{b}(1), \ldots, \hat{b}(B-1)$, where B is the size of original transport block information with attached CRC bits.

5. *CRC check*.
   CRC check is used to extract the CRC bits in UCI/DCI bits after the information transmitted in 5G NR control channels. Then the extracted CRC bits will be checked with the original CRC bits attached to control information bits before transmitted.

1.3.4 Polar Decoding Algorithms

Two primary polar decoding methods are SC decoder and BP decoder. Unlike SC decoder, BP decoder doesn't have inter-bit dependence, preventing error propagation and avoiding intermediate hard decisions. It updates LLR values iteratively through right-to-left and left-to-right iterations using LDPC-like update functions. BP decoder supports parallel processing, enhancing throughput for high-speed applications, while SC decoder and its variants have serial decoding characteristics, making parallelization impossible [18] . Moreover, polar decoding can employ various algorithms, including SC decoding and SCL decoding. Polar decoders face challenges in hardware implementation compared to encoders due to complexities: they work with bit probabilities, consider all possible permutations of information blocks, and process each block multiple times for error correction. This leads to higher latency, hardware usage, and power consumption.

(1) ***Successive Cancellation Decoding***:

The first polar code decoding method, known as successive cancellation (SC), decodes bits one by one, using previous estimations to help determine new ones [85]. SC builds on decoded bits sequentially but suffers from inter-bit dependence and error spread. While it doesn't perform as well as other decoders alone, it shows promise for list decoding because of its hierarchical structure. Polar codes achieve the Shannon capacity under SC decoding. The computation method for SC is akin to that of LDPC codes, using log-likelihood ratios (LLRs) to estimate bit likelihoods. SC decoding algorithms utilize the log-likelihood function, which can be recursively computed due to the recursive nature of the channel transition function.

Let

$$\lambda_{\mathsf{N}}^{(i)}\left(\mathrm{y}_0^{\mathsf{N}-1}, \hat{\mathrm{u}}_0^{i-1}\right) = \frac{\mathrm{W}_{\mathsf{N}}^{(i)}\left(\mathrm{y}_0^{\mathsf{N}-1}, \hat{\mathrm{u}}_0^{i-1} \mid \mathrm{u}_i = 0\right)}{\mathrm{W}_{\mathsf{N}}^{(i)}\left(\mathrm{y}_0^{\mathsf{N}-1}, \hat{\mathrm{u}}_0^{i-1} \mid \mathrm{u}_i = 1\right)}, \tag{2.40}$$

and let

$$\mathscr{L}_{\mathsf{N}}^{(i)}\left(\mathrm{y}_0^{\mathsf{N}-1}, \hat{\mathrm{u}}_0^{i-1}\right) = \log \lambda_{\mathsf{N}}^{(i)}\left(\mathrm{y}_0^{\mathsf{N}-1}, \hat{\mathrm{u}}_0^{i-1}\right). \tag{2.41}$$

This likelihood ratio can be used to estimate the value of the bit u_i using the function h_i defined as

$$\mathrm{h}_i\left(\mathrm{y}_0^{\mathsf{N}-1}, \hat{\mathrm{u}}_0^{i-1}\right) = \begin{cases} 0 & \text{if } \mathscr{L}_{\mathsf{N}}^{(i)}\left(\mathrm{y}_0^{\mathsf{N}-1}, \hat{\mathrm{u}}_0^{i-1}\right) \geq 0, \\ 1 & \text{otherwise .} \end{cases} \tag{2.42}$$

For decoding, if $i \in \mathcal{A}^c$ (i.e., the set of frozen bits), then the decoded value is the frozen bit value. Otherwise, it is determined by the h function:

$$\hat{\mathrm{u}}_i = \begin{cases} \mathrm{u}_i & i \in \mathcal{A}^c, \\ \mathrm{h}_i\left(\mathrm{y}_0^{\mathsf{N}-1}, \hat{\mathrm{u}}_0^{i-1}\right) & i \in \mathcal{A} . \end{cases} \tag{2.43}$$

Note that $\hat{\mathrm{u}}^i$ depends on the previously estimated values $\hat{\mathrm{u}}_0^{i-1}$. This is the essence of SC: bits are estimated in order $\hat{\mathrm{u}}_0, \hat{\mathrm{u}}_1, \hat{\mathrm{u}}_2, \ldots, \hat{\mathrm{u}}_{\mathsf{N}-1}$, with the estimate $\hat{\mathrm{u}}^i$ being based upon previously determined bits. Under the polarization idea, since the polarized channels used are assumed to be good, each of the previously determined bits $\hat{\mathrm{u}}_0^{i-1}$ are assumed to be good [85].

Moreover the SC decoding principle for polar codes requires only two clarifications: the first is the probability transfer formula on the unit factor graph, and the second is the recursive order. As Figure 2.9 shows the graph of the unit factor of the polar code, on which there are 8 values, $\mathscr{L}_1, \mathscr{L}_2, \mathscr{L}_3, \mathscr{L}_4$ for the LLR values passed to the left and $\mathscr{B}_1, \mathscr{B}_2, \mathscr{B}_3, \mathscr{B}_4$ for the hard bit information passed to the right.

The transmission equation is

$$\mathscr{L}_3 = \mathscr{L}_1 \boxplus \mathscr{L}_2, \tag{2.44}$$

$$\texttt{where : } \mathrm{a} \boxplus \mathrm{b} = 2\tanh^{-1}\left[\tanh\left(\frac{a}{2}\right)\tanh\left(\frac{b}{2}\right)\right]. \tag{2.45}$$

$$\mathscr{L}_4 = \begin{cases} \mathscr{L}_2 + \mathscr{L}_1 & \texttt{if } \mathscr{B}_3 = 0 . \\ \mathscr{L}_2 - \mathscr{L}_1 & \texttt{if } \mathscr{B}_3 = 1 . \end{cases} \tag{2.46}$$

$$\begin{aligned} \mathscr{B}_1 &= \mathscr{B}_3 \oplus \mathscr{B}_4, \\ \mathscr{B}_2 &= \mathscr{B}_4. \end{aligned} \tag{2.47}$$

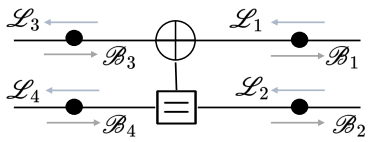


Figure 2.9: Factor Graph of the SC decoding unit.

In what follows, we'll describe how the recursive process works using the factor graph shown in Figure 2.10. First, the decoder needs to find the value called LLR at point 1. To do this, it has to know the LLR values at points 2 and 3. Similarly, to find the LLR at point 2, it needs the LLR values at points 4 and 5, and for point 3, it needs the LLR values at points 6 and 7. However, the LLR values at points 4, 5, 6, and 7 can be figured out directly from the LLR values sent through the channel.

Then, we use equation (2.44) to calculate the LLR value at point 8 using the LLR value at point 1. Since point 8 is in the bottom left of the *unit factor graph*, we already know the LLR values of points 2 and 3, so we don't need to do any more recursion. Once we find the LLR value for point 8, we make a firm decision. After finding the LLR value for the bottom left point, we know the bit decision should move to the right. In other words, we figure out the bit values for points 2 and 3 using equations (2.45) and (2.46). Now that points 2 and 3 are in the top left of the unit factor graph, we don't need to pass hard decision bit values to the right anymore. Next, to find the LLR value at point 9, we first need to calculate the LLR values at points 10 and 11. Since points 10 and 11 are in the bottom left, we can use equation (2.44) along with the bit decisions of points 2 and 3. We don't need any more downward recursion. Similarly, we find the LLR value at point 12 using equation (2.44) and the bit decision of point 9, then we make a decision. As point 12 is in the bottom left, we move the hard decision bit value to the right. At this stage, we compute the binary values of points 10 and 11 using equations (2.45) and (2.46), and since they are both in the bottom left, we continue moving binary values to the right for points 4, 5, 6, and 7. Once points 4, 5, 6, and 7 are all in the top left, we stop moving binary values to the right. We repeat this process until the SC decoder has found all LLR values and assigned binary values.

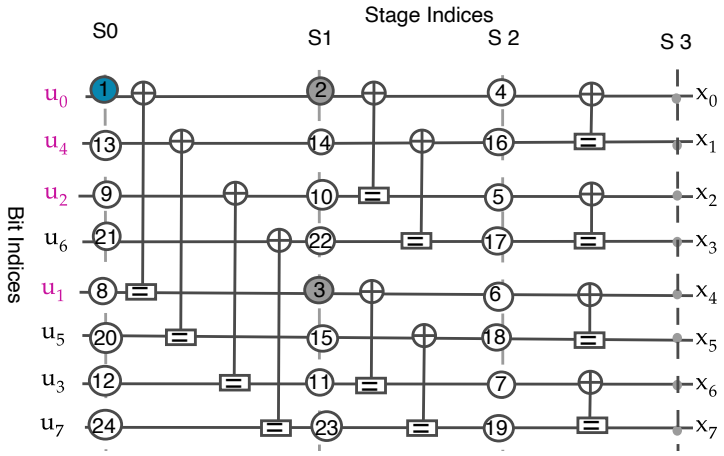


Figure 2.10: Illustrative instance of a recursive order.

Remark 2.2. Despite the recursive SC method, the SC method can also be calculated using the node labelling method. To program the LLR recursion process described above and the process of passing the hard decision bit value to the right, the points shown in Figure 2.10 must be labelled so that they can be programmed in a certain order.

(2) ***Successive Cancellation List Decoding*** :

The SCL decoder was introduced as an extension of the SC decoder. Rather than sequentially computing hard decisions for each bit, it bifurcates into two parallel SC decoders at every decision stage, with each branch maintaining its path metric continuously updated for each path. It's demonstrated that a list size of 32 is nearly sufficient to reach the ML bound [18].

What is needed is a path metric, computing the likelihood along the entire path of bits. This path metric is established in the following theorem.

Theorem 2.1. [52, Theorem 1].
For a path ℓ having bits $\hat{\mathrm{u}}_0(\ell), \hat{\mathrm{u}}_1(\ell), \ldots, \hat{\mathrm{u}}_i(\ell)$ and bit number $i \in 0, 1, \ldots, \mathrm{N}-1$, define the path metric as

$$\mathsf{PM}_{\ell}^{(i)} = \sum_{j=0}^{i} \ln\left(1 + \exp\left[-\left(1 - 2\hat{\mathrm{u}}_j(\ell)\right)\mathscr{L}_{\mathsf{N}}^{(j)}[\ell]\right]\right) \tag{2.48}$$

where

$$\mathscr{L}_{\mathsf{N}}^{(i)}[\ell] = \ln\left(\frac{\mathrm{W}_{\mathsf{N}}^{(i)}\left(\mathrm{y}_0^{\mathsf{N}-1}, \hat{\mathrm{u}}_0^{i-1}[\ell] \mid 0\right)}{\mathrm{W}_{\mathsf{N}}^{(i)}\left(\mathrm{y}_0^{\mathsf{N}-1}, \hat{\mathrm{u}}_0^{i-1}[\ell] \mid 1\right)}\right) \tag{2.49}$$

is the LLR of the bit u_i given the channel output $\mathrm{y}_0^{\mathsf{N}-1}$ and the past trajectory of the path $\hat{u}_0^{i-1}[\ell]$. However, the path metric is computed using LLRs, in a numerically stable way[85].

Furthermore, CRC-aided SCL, an extension of the SCL decoder, incorporates a high-rate CRC code appended to the polar code. This addition facilitates the selection of the correct codeword from the final list of paths. It has been observed that in instances where an SCL decoder fails, the correct codeword remains within the list. Hence, the CRC serves as a validation check for each candidate codeword in the list. Polar decoding via BP and SCAN is beyond the scope of this manuscript. Interested readers are encouraged to refer to [69, 72] for BP decoding and [58] for SCAN decoding.

2 Modulation Sets

5G NR supports *quadrature phase shift keying* (QPSK), *quadrature amplitude modulation* (16-QAM), 64-QAM and 256-QAM for uplink and downlink [133, 134]. This is the same as for LTE. In addition, 3GPP supports $\pi/2$-BPSK uplink modulation in *cyclic prefix* (CP)-OFDM or *discrete Fourier transform-spread* (DFT-s) OFDM + CP transmissions to further reduce the *peak-to-average power ratio* (PAPR).

2.1 BPSK / $\pi/2$ - BPSK

$\pi/2$ - BPSK uses two sets of BPSK constellations that are shifted shifted by $\pi/2$ radians (90°). The constellation sets are selected depending on the position of the bits in the input sequence. In case of BPSK and $\pi/2$ - BPSK modulation, bit $b(i)$ is mapped to complex-valued modulation symbol $d(i)$ according to

$$d(i) = \begin{cases} \frac{1}{\sqrt{2}}[(1-2b(i)) + j(1-2b(i))], & \texttt{if BPSK.} \\ \frac{1}{\sqrt{2}}\exp(j\frac{\pi}{2}(i \bmod 2))[(1-2b(i)) + j(1-2b(i))], & \texttt{if } \pi/2\texttt{–BPSK.} \end{cases} \tag{2.50}$$

$b(i)$ = input bits; i = position or index of input bits; $d(i)$ = mapped bits (constellation points). From the waveform, one may note that $\pi/2$ - BPSK has more phase transitions than BPSK.

2.2 QPSK / $\pi/4$-QPSK

A QPSK signal essentially combines two orthogonally modulated BPSK signals. Its performance in terms of bit error rate is virtually the same as that of conventional BPSK. It has several variants, including $\pi/4$-QPSK modulation, which is characterised by the fact that the signalling points of the modulated signals are chosen from two QPSK constellations that are just $\pi/4$ radians (45°) apart, thus preserving the constant envelope property better than QPSK modulation. Thus, pairs of bits, $b(2i)$, $b(2i+1)$) , are mapped to complex-valued modulation symbols $d(i)$.

$$d(i) = \begin{cases} \frac{1}{\sqrt{2}}[(1-2b(2i)) + (1-2b(2i+1))], & \texttt{if QPSK.} \\ \frac{1}{\sqrt{2}}\exp(j\frac{\pi}{4}(i \bmod 2))[(1-2b(2i)) + j(1-2b(2i+1))], & \texttt{if } \pi/4\texttt{–QPSK.} \end{cases} \tag{2.51}$$

2.3 M-ary Quadrature Amplitude Modulation (M-QAM)

In this instance, the information bits are encoded as variations in the amplitude and phase of the signal.

2.3.1 16 QAM

For 16 QAM modulation set, quadruplets of bits, $b(4i)$, $b(4i+1)$, $b(4i+2)$, $b(4i+3)$, are mapped to complex-valued modulation symbols $d(i)$ as

$$d(i) = \frac{1}{\sqrt{10}}\{(1-2b(4i))[2-(1-2b(4i+2))] + j(1-2b(4i+1))[2-(1-2b(4i+3))]\}. \tag{2.52}$$

2.3.2 64 QAM

For 64 QAM modulation set, hextuplets of bits, $b(6i)$, $b(6i+1)$, $b(6i+2)$, $b(6i+3)$, $b(6i+4)$, $b(6i+5)$, are mapped to complex-valued modulation symbols $d(i)$ such that

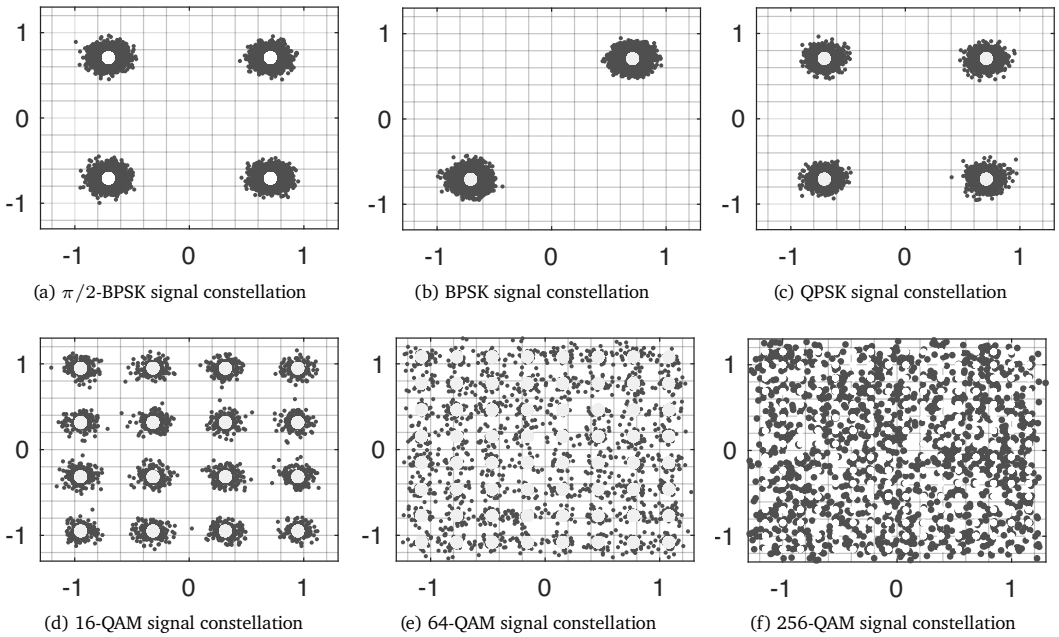
$$\begin{aligned} d(i) =& \frac{1}{\sqrt{42}}\{(1-2b(6i))[4-(1-2b(6i+2))[2-(1-2b(6i+4))]] \\ &+ j(1-2b(6i+1))[4-(1-2b(6i+3))[2-(1-2b(6i+5))]]\}. \end{aligned} \tag{2.53}$$

2.3.3 256 QAM

For 256 QAM modulation set, octuplets of bits, b(8i),$b(8i+1)$, $b(8i+2)$, $b(8i+3)$, $b(8i+4)$, $b(8i+5)$, $b(8i+6)$, $b(8i+7)$, are mapped to complex-valued modulation symbols $d(i)$ as follows.

$$\begin{aligned} d(i) =& \frac{1}{\sqrt{170}}\{(1-2b(8i))[8-(1-2b(8i+2))[4-(1-2b(8i+4))[2-(1-2b(8i+6))]]]+ \\ & j(1-2b(8i+1))[8-(1-2b(8i+3))[4-(1-2b(8i+5))[2-(1-2b(8i+7))]]]\}. \end{aligned} \tag{2.54}$$

Figure 2.11 shows a schematic representation of some constellation sets for common linear modulations along with Gray mapping. In practical terms, a bit mapping using Gray coding means that every two adjacent m-bit symbols differ by a single bit. Due to better *bit error rate* performance, Gray mapping under QPSK modulation will be considered and used for most simulations in this work, unless otherwise stated.



(a) $\pi/2$-BPSK signal constellation (b) BPSK signal constellation (c) QPSK signal constellation

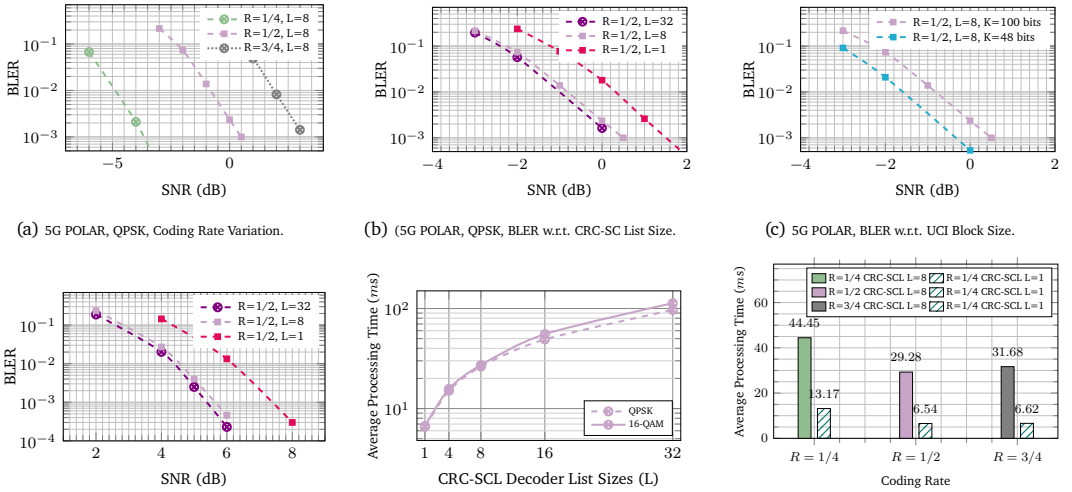
(d) 16-QAM signal constellation (e) 64-QAM signal constellation (f) 256-QAM signal constellation

Figure 2.11: Gray Labelling modulation constellation sets at noise-free and SNR=-20 dB.

3 5G Coded Modulation Performance analysis

Since we focus on the short block regime, i.e., blocks ranging from 20 to 100 bits, simulations are conducted using a 100-bit block (including TB/UCI and CRC) with various coding rates from low to high. The study considers both a low-order modulation scheme (QPSK) and a high-order modulation scheme (16-QAM), in combination with polar and LDPC coding.

Starting with the 5G polar coded modulation configuration, Figure 2.12a shows the reliability performance in terms of BLER as a function of the code rate. It can be observed that lower code rates yield better performance. In Figures 2.12b and 2.12d, at a fixed code rate, the performance of the CRC-aided SCL decoder is evaluated for different list sizes: L=1 (equivalent to the baseline CRC-SC), L=8, and L=32. The results indicate that performance improves with increasing list size, and as L approaches 32, the decoding becomes more refined.
In Figure 2.12c, at a fixed code rate, the impact of varying the input block size is analyzed. Smaller input UCI sizes lead to better decoding performance.
Figure 2.12f illustrates the evolution of decoding complexity, measured as the average processing time at a fixed SNR point over 100 Monte Carlo trials. The figure evaluates the duration of CRC-aided SCL decoding as a function of the list size L. It is observed that decoding complexity grows exponentially with L for both low- and high-order modulation schemes, namely QPSK and 16-QAM.
Moreover, Figure 2.12f demonstrates that in SCL decoding, complexity is influenced not only by the total number of coded bits (codeword size) but also by the number of active information bits. Higher code rates activate more bits, generating additional paths within the list and consequently slowing down the decoder (e.g., $R = 1/2$ versus $R = 3/4$). This occurs even though, intuitively, a lower total number of bits might suggest lower complexity. Among the configurations tested, $R = 1/2$ appears to offer the optimal trade-off, as it results in the lowest complexity due to the balance between active information bits K and frozen bits $(N - K)$.
Overall, the CRC-aided SCL decoder is widely recognized for its superior error-correction capability compared to state-of-the-art decoders. The overall performance of the SCL decoder improves as the list size (L) increases. Indeed, increasing L in the SCL decoder enhances the BLER performance; however, this improvement comes at the cost of higher decoding complexity and increased computational resource requirements.



(a) 5G POLAR, QPSK, Coding Rate Variation. (b) (5G POLAR, QPSK, BLER w.r.t. CRC-SC List Size. (c) 5G POLAR, BLER w.r.t. UCI Block Size.

(d) 5G POLAR, 16-QAM, BLER w.r.t. CRC-SC List Size. (e) 5G POLAR, Complexity w.r.t. List Size. (f) 5G POLAR, Complexity w.r.t. Coding Rate (Codeword length).

Figure 2.12: Block error rate (BLER) and average processing complexity analysis. **polar coded modulation parameters:** UCI bits(89 bits) + CRC(11 bits) = 100 bits, coding rate R={1/4, 1/2, 3/4}, modulation schemes= QPSK, 16QAM. **Decoders:** CA-SCL, L={1, 8, 32}. **Channel configurations:** Uplink transmission, perfect CSI, 1×4 SIMO 3GPP TDL-C NLOS (long delay spread = 300 ns, Urban Macro, sampling rate $f_s = 30.72$ MHz).

In the same vein, Figure 2.13 highlights both the reliability and complexity performance of the 5G LDPC coded modulation configuration.
Similar to the polar coded modulation configuration, Figure 2.13a presents the reliability performance in terms of BLER as a function of the code rate. It can be observed that lower code rates provide better error-correction performance.
Figures 2.13b and 2.13d evaluate the performance of the LBP, baseline BP, and OMS decoders at a fixed code rate, under QPSK and 16-QAM modulation configurations. Two scenarios are considered: first, when the decoder runs until the maximum number of iterations is reached; and second, when early stopping criteria (ESC) are employed to enable faster convergence once successful decoding is achieved. The results show that the LBP decoder consistently achieves slightly better performance than the baseline BP decoder for the same number of iterations. Furthermore, when ESC is applied, in cases where LBP successfully decodes a frame in a single iteration, the BP decoder typically requires two iterations. This indicates that, in terms of convergence speed, LBP can be up to twice as fast as BP.

From a structural perspective, the LBP decoder outperforms its standard BP counterpart due to its ability to more efficiently exploit the inherent structure of LDPC codes. In practice, LDPC codes are commonly constructed with a layered structure, meaning that the parity-check matrix can be partitioned into subsets (layers) of check nodes associated with specific variable nodes. The LBP algorithm leverages this structure by performing decoding sequentially, layer by layer. This approach enables more efficient message passing compared to the conventional "flooding" BP schedule, in which all variable-node and check-node messages are updated simultaneously within each iteration. By updating messages immediately within each processed layer, LBP accelerates the propagation of reliability information across the Tanner graph, thereby improving convergence speed. Moreover, LBP generally achieves lower effective computational complexity. Since message updates are confined to one layer at a time, intermediate information is reused within the same iteration, reducing the number of required global iterations to reach convergence. This results in improved convergence behavior and enhanced error-correction performance compared to standard flooding-based BP decoding. Such advantages are particularly significant under challenging channel conditions or at low-to-moderate SNR regions. Hence, the layered structure and optimized message-passing schedule make LBP a preferred decoding strategy for LDPC codes in many practical implementations, offering improved reliability, faster convergence, and reduced effective complexity. Nevertheless, the relative performance of LBP and BP decoding remains dependent on several factors, including the code rate, the operating signal-to-noise ratio (SNR), the structural properties of the LDPC code, and the available computational resources.

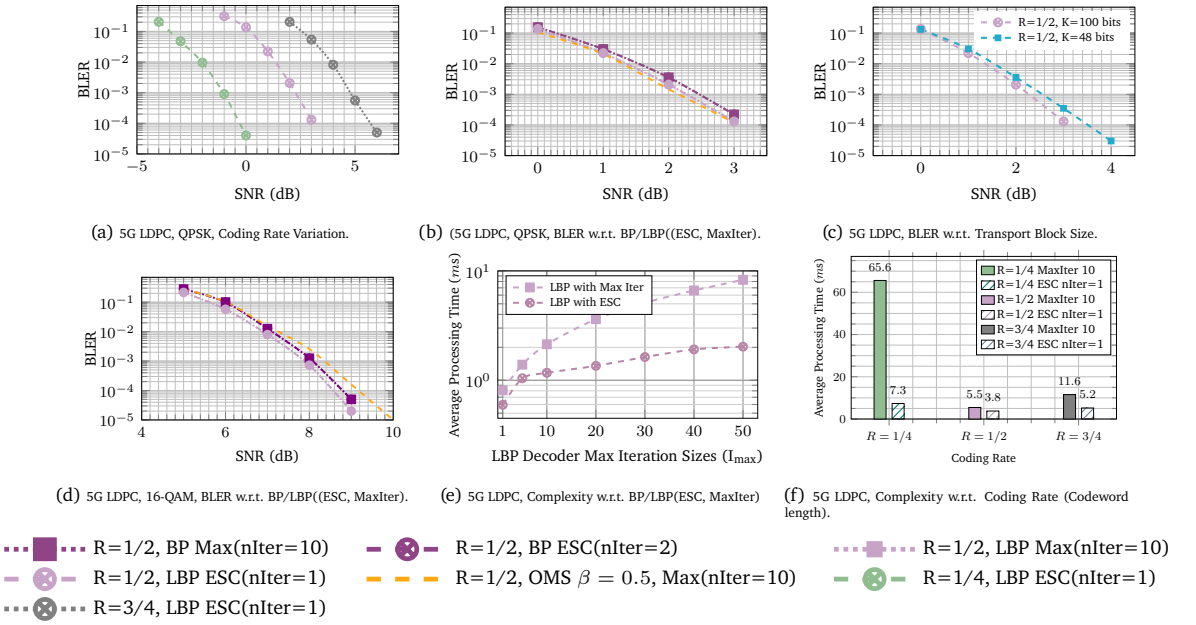
Subsequently, In Figure 2.13c, the impact of varying the input block size is analyzed at a fixed code rate. In principle, smaller transport block sizes are expected to yield better decoding performance due to the reduced codeword length and potentially improved reliability. However, the LDPC code exhibits a different behavior. Specifically, the performance obtained with TBS+CRC = 100 bits is slightly better than that with TBS+CRC = 48 bits. This behavior can be explained by the fact that the LDPC codes defined in the 3GPP standard are primarily optimized for moderate to large block lengths. For short block sizes, the encoder and decoder may not operate under their most favorable structural conditions, leading to suboptimal or unexpected performance behavior. In contrast, polar codes are known to perform well in the short-packet regime. Therefore, the observed discrepancy can be attributed to the intrinsic design characteristics of LDPC codes compared to polar codes, particularly in the context of short block lengths.

Figure 2.13e illustrates the evolution of decoding complexity, measured in terms of the average processing time at a fixed SNR over 100 Monte Carlo trials. The figure evaluates the LBP decoding time as a function of the maximum number of iterations (Imax). As expected, the decoding complexity increases significantly with Imax. However, when early stopping criteria (ESC) are applied, the decoder converges after a limited number of iterations that is typically much smaller than $\mathrm{I}_{\max}$, regardless of its preset value. This results in a substantial reduction in decoding time.

Furthermore, Figure 2.13f shows that the complexity of LBP decoding is influenced not only by the maximum number of iterations but also by the code structure, particularly the total number of coded bits (N), thus the number of information bits (K), and the number of parity bits ($N - K$). Higher code rates activate a larger proportion of information bits, increasing bit-node processing during each iteration and consequently slowing down the decoder (e.g., $R = 3/4$ compared to $R = 1/2$). This occurs even though one might intuitively expect fewer total bits to imply lower complexity. Conversely, lower code rates activate a larger number of parity bits, which increases check-node processing within each iteration and also leads to higher decoding complexity (e.g., $R = 1/4$ compared to $R = 1/2$). Among the evaluated configurations, the code rate $R = 1/2$ appears to provide the most favorable trade-off, as it achieves the lowest decoding complexity due to the balanced proportion between information bits (K) and parity bits ($N - K$).

As a result, the LBP decoder with ESC and a moderate code rate appears to represent a sweet-spot solution, effectively balancing reliability and computational complexity.

As a whole, considering the aforementioned analysis, using a higher-order modulation would require a detector compatible with the transmitter's modulation, increasing receiver complexity alongside channel estimation and decoding stages. Therefore, for scenarios involving short packet transmission, a simpler modulation like QPSK may suffice. Thus, combining a soft QPSK-based demodulator with the efficient ESC-based LBP decoder for the LDPC coded modulation and CA-SCL decoder (with a limited number of lists) for polar coded modulation configurations seems to offer a suitable performance/complexity trade-off in short packet reception.



(a) 5G LDPC, QPSK, Coding Rate Variation. (b) (5G LDPC, QPSK, BLER w.r.t. BP/LBP((ESC, MaxIter). (c) 5G LDPC, BLER w.r.t. Transport Block Size.

(d) 5G LDPC, 16-QAM, BLER w.r.t. BP/LBP((ESC, MaxIter). (e) 5G LDPC, Complexity w.r.t. BP/LBP(ESC, MaxIter) (f) 5G LDPC, Complexity w.r.t. Coding Rate (Codeword length).

Figure 2.13: Block error rate (BLER) and average processing complexity analysis. **LDPC coded modulation parameters:** TB (84 bits) + CRC(16 bits) = 100 bits, coding rate R={1/4, 1/2, 3/4}, modulation schemes= QPSK, 16QAM. **Decoders:** BP, LBP with Early Stop Criteria (ESC),0MS β=0.5, MaxIter=10. **Channel configurations:** Uplink transmission, perfect CSI, 1×4 SIMO 3GPP TDL-C NLOS (long delay spread = 300 ns, sampling rate $f_s = 30.72$ MHz).

4 Summary

This chapter has provided backgrounds and introduced preliminary knowledge of LDPC and POLAR coded modulation, emphasizing their relevance to the 5G NR standard. The design process underlying these channel codes was highlighted to understand their foundations and fundamentals from the encoder to decoder standpoints. Additionally, a literature overview was provided to establish the state of the art of current knowledge in the field. Finally, the modulation schemes commonly used in 5G NR communication channels were explored, with discussions on their potential incorporation with 5G Polar and 5G LDPC coding for URLLCs. The following chapters will deal with the main contributions that arise from this work, namely the design of advanced BICM receivers.

Chapter 3

Receiver Design (Part I)

Contents

1 Relevant prior of art

Zehavi [86] proposed bit-interleaved coded modulation (BICM) as a pragmatic approach to coded modulation. Its basic principle is the ability of an interleaving permutation to separate an underlying binary code from an arbitrary higher-order modulation [87]. Per-bit log-likelihood ratios are used to convey soft metrics from the demodulator to the decoder in order to reduce information loss. Thus, this fundamental observation spurred interest in BICM. Later, Caire *et al.* [88] conducted a comprehensive theory of bit-interleaved coded modulation and offering guidelines for its design, including both coherent and non-coherent detection. Afterwards, numerous research inquiries have been directed towards the design of low-complexity receivers for BICM systems, with a particular focus on low-dimensional and high-dimensional MIMO systems, but primarily restricted to coherent communication. BICM MIMO OFDM emerges as an attractive prospect for future wireless networks, wherein MIMO enhances spectrum efficiency, OFDM reduces equalization complexity, and BICM provides reliable coded-modulations. Among which, BICM MIMO OFDM becomes an attractive prospect for future wireless networks, where MIMO improves spectrum efficiency, OFDM reduces equalization complexity and BICM provides reliable coded modulations. Accordingly, in the early years of the 21st century, significant progress was made in the design of ML receivers adapted to MIMO systems [89–91]. Specifically, Hochwald and Marzetta [91] proposed unitary space-time modulation, which eliminates the need for receiver knowledge of propagation coefficients, showcasing its efficacy in achieving high performance even without perfect channel state information.

Advances in coded modulation schemes have prompted significant interest and exploration in recent literature showcasing advancements in performance, complexity reduction, and applicability across diverse communication scenarios. In reviewing the literature, Xie *et al.* [92] propose a polar coded modulation with physical network

coding, achieving significant performance gains over conventional BICM systems. Meanwhile, Wang *et al.* [93] discuss adaptive modulation and coding in 5G systems, emphasizing the importance of spectrum efficiency through channel state information feedback. Later, Mu *et al.* [94] introduce BICM with iterative SCL decoding, showing improved performance and reduced complexity. Conversely, Tian *et al.* [95] present joint SC decoding for *bit-interleaved polar coded modulation*(BIPCM), enhancing existing BIPCM schemes with improved decoding algorithms. Yuan *et al.* [96] propose polar-coded non-coherent communication, demonstrating competitive performance in non-coherent systems compared to coherent ones. Moreover, Mahdavifar *et al.* [97] introduce polar coding for BICM, highlighting the efficiency of polar codes over multi-channels. Shieh *et al.* [98] consider systematic polar coded modulation for *informed* receivers, addressing coding for receivers with partial side information and achieving significant gains. Later, Wan *et al.* [99] discuss *short-block length*(SBL) *polar-coded modulation*(PCM) for the Relay Channel, showing substantial improvements in performance for relayed communications. Dai *et al.* [100] propose polar-coded spatial modulation, integrating polar coding and spatial modulation, for enhanced performance in MIMO systems. Meanwhile, Li *et al.* [101] present an URLLC non-binary polar coded *sparse code multiple access* (SCMA) scheme, offering solutions for latency and reliability issues in SCMA systems. Furthermore, Yuan *et al.* [102] introduce two-Level *ldpc-coded modulation* (LCM) with probabilistic shaping, demonstrating higher spectral efficiency in future communications. Choi [103] investigates MIMO-BICM iterative receivers, focusing on channel estimation and decoding techniques for improved performance, while Chiu [104] discusses the analysis and design of PCM, proposing novel design methods to enhance performance under SCL decoding. Meanwhile, Xu *et al.* [105] analyzed the limitations of coding and modulation schemes in *long-term evolution* (LTE)-based 5G terrestrial broadcast systems, proposing enhancements using LDPC codes and non-uniform constellations for significant performance gains without added complexity, while Chen *et al.* [106] explored the use of polar codes in non-coherent systems, introducing innovative detection and decoding techniques to reduce performance discrepancies and enhance practical applicability compared to coherent systems.

Recently, there has been a growing interest in applying deep learning to coded modulation systems, with a particular interests in receiver design aspects, promising avenues for enhancing the efficiency and reliability of wireless communication systems. Carrera *et al.* [107] introduced an *extreme learning machine* (ELM)-based receiver for multi-user massive MIMO systems, offering higher spectral efficiency and smaller bit error rates compared to conventional linear receivers. Subsequently, Cammerer *et al.* [108] presented a neural network (NN)-based MU-MIMO receiver compatible with 5G NR standards, demonstrating close-to-optimal performance with significantly lower computational complexity. Additionally, Li *et al.* [109] utilized deep learning to improve the efficiency of a BICM-ID system by learning bit posterior probabilities, showcasing promising results under various channel conditions and modulation modes. Meanwhile, Doan [110] addressed the challenges of decoding short linear block codes by proposing novel techniques tailored to SC decoding algorithms to reduce decoding latency and complexity while maintaining competitive error-correction performance and by leveraging machine learning algorithms, the proposed decoders offer efficient optimization and demonstrate superior performance compared to state-of-the-art approaches.

Upon revisiting the core of this investigation, considerable research efforts have been dedicated to various facets, including the design of signal codes, and receivers in the realm of reliable low-latency short data communication, especially with the emergence of 5G and 6G wireless standards. Accordingly, Lee *et al.* [111] emphasized the need for packet structure and receiver designs suitable for ultra-short packet transmissions, essential for meeting the stringent latency requirements. Meanwhile, Sy and Knopp [112] focused on enhancing the performance of short block transmission systems by novel receiver structures paired with advanced joint estimation detection technique. They additionally highlighted the importance of incorporating training information and adaptive power adjustment strategies to optimize performance. Meanwhile, they explored in [113] the use of block/segment coding to mitigate complexity issues associated with long bit length codes within the short block regime. Subsequently, authors in [61, 114] delved into decoding techniques for ultra-reliable low-latency communications (URLLC) for SBL codes, highlighting the necessity of efficient decoders capable of achieving superior error-rate performance while maintaining low complexity. They discussed various decoding strategies and proposed algorithms to address latency concerns in polar codes, a popular choice for URLLC applications. Moreover, Shirvanimoghaddam *et al.* [115] and Yue *et al.* [116] provided comprehensive reviews of channel coding techniques within SBL codes for URLLC, emphasizing the need for error-rate performance, low latency, and rate compatibility. They compared different decoding techniques in terms of performance and complexity and identified potential research directions in this area.

The application of joint estimation and detection techniques to improve receiver performance has been largely

studied, although in the area of short packet reception there are still limits that need to be explored. Jalali *et al.* [117] proposed a novel algorithm for joint detection and decoding of polar-coded 5G control channels, aiming to overcome practical obstacles such as CSI errors and interference. Huang *et al.* [118] introduced the *sparsely interleaved estimation and decoding* (SIED) algorithm for bit-interleaved coded modulation (BICM) over correlated fading channels, effectively preventing error propagation and improving decoding results. Subsequently, Wu *et al.* [119] presented a CRC-aided joint iterative detection and SCL decoding receiver for polar coded SCMA systems, achieving better error rate performance and lower complexity compared to existing methods, while Sun *et al.* [71] proposed an LDPC-coded SCMA receiver with multi-user iterative detection and decoding, offering a low-complexity realization and high throughput for SCMA systems. Meanwhile, Jiao *et al.* [121] introduce a joint channel estimation and decoding scheme for polar coded SCMA systems over fading channels in the wake of enhancing 5G communications for mMTC within the IoT. Their proposed joint estimation-decoding scheme relies on *sparse Bayesian learning* for initial CSI measurements, iteratively refines channel estimation and detection for enhanced system performance. In another direction, Marey *et al.* [122] proposed an iterative receiver design for uplink OFDMA cooperative systems, addressing the joint data detection and channel estimation problem. Yuan *et al.* [123] investigated the design of *orthogonal time frequency space* (OTFS) detectors based on approximate message passing, demonstrating superior performance and significant improvements over existing detectors. Conversely, Li *et al.* [124] explored an efficient iterative receiver for BICM according to the *digital video broadcasting – second generation terrestrial* (DVB-T2) standard, highlighting the potential of iterative processing for practical applications, showing the merits of the iterative process in achieving progressive performance. In recent advancements, the utilization of *superimposed pilots* (SIP) in wideband multiuser MIMO uplink systems has been explored by Qian *et al.* [125] to mitigate the overhead associated with traditional pilot schemes and enhance spectral efficiency. The challenge lies in effectively decoupling the superimposed data and pilot signals, which share the same power domain. They propose an iterative joint channel estimation and detection approach, leveraging soft-symbol information propagation to enable pilot and data interference cancellation, demonstrating significant improvements in detection accuracy and overall system throughput, positioning it as a promising solution in the optimization of MIMO systems.

Another important aspect relating to the transmission of short packets surrounds the state-of-the-art converse and achievability bounds for both coherent and non-coherent communications [126–131]. This includes finite block length information theoretic tools utilized in assessing coding schemes. For the derivation of outer bounds (converse), reference is in particular made to the *metaconverse* (MC) theorem [127, Th. 28]. Additionally, for the inner bound (achievability), a relaxed version of the *random coding union* (RCU) bound [127, Th. 5], widely recognized as the RCUs bound is provided by Martinez and i Fàbregas [130, Th. 1].

This work distinguishes itself from prior literature by introducing novel enhanced receiver designs tailored to scenarios with imperfect *channel state information* (CSI) over various channel conditions. The proposed receivers are specifically designed to effectively support the reception of short data packets in beyond 5G/6G signaling scenarios by evaluating their performance over 5G short block channels using both Polar and LDPC coded modulation formats.

2 General Framework

2.1 Bit-Interleaved coded Modulation

2.1.1 Bit-Interleaved Polar-coded Modulation

In this instance of BIPCM, we are dealing with the CRC-aided polar coding scheme. One can refer to Chapter 2 for details of the encoding/decoding procedure in use. The transmission process is straightforward and complies with the 3GPP standard specifications [132]. Since CA-SCL decoding is capable of outperforming Turbo or LDPC codes, it will be retained as the ideal candidate decoder in the remainder of this chapter.

2.1.2 Bit-Interleaved LDPC-coded Modulation

The transmission procedure of BILCM is almost identical to that described with BIPCM. One can refer to Chapter 3 for details of the encoding/decoding procedure in use. To strike a better balance between performance and complexity, the layered belief propagation (LBP) algorithm, which stands out as a promising approach for URLLC due to its ability to speed up convergence times, is considered.

Furthermore, Figure 3.1 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* (TB) CRC, segmenting code blocks with additional

CRC attachment, channel encoding, rate matching, code block concatenation, modulation and resource mapping. It is important to emphasize that the receiving chain simply follows the reverse flow of the transmitting chain.

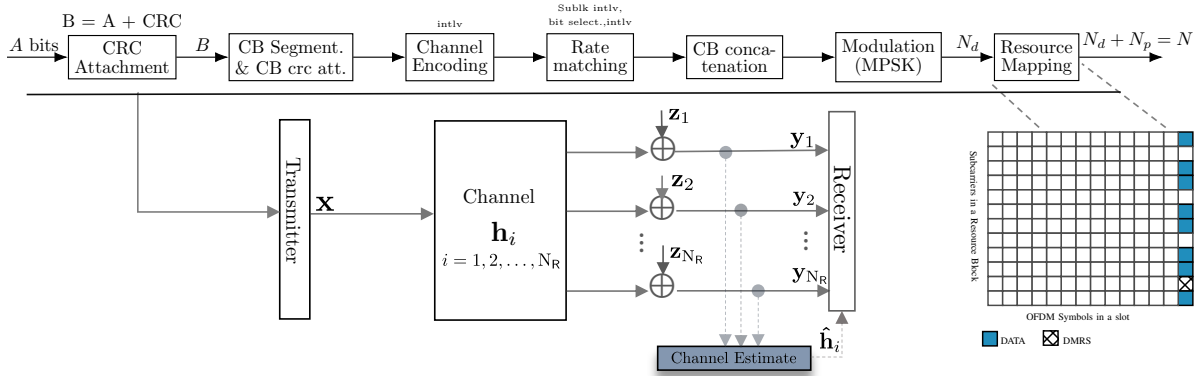
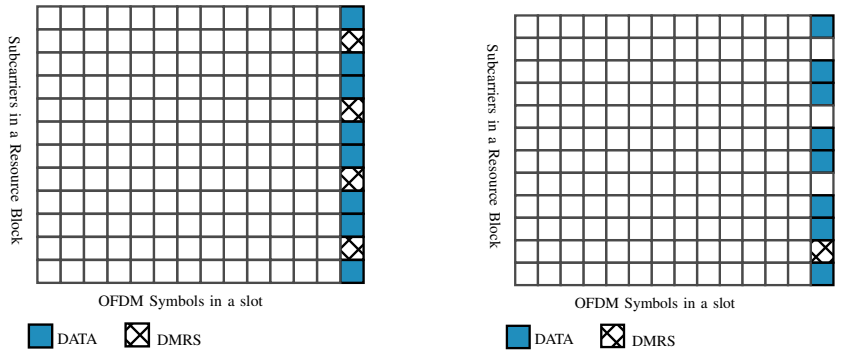


Figure 3.1: Bit-Interleaved Polar/LDPC coded Modulation (BIPCM/BILPCM) : Transmitter end.

2.2 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 3.2, the resource mapping process in use is in the same spirit as the 3GPP PUCCH format 2 transmission [134] for payload size of up to 11 bits.



(a) Moderate DMRS density (e.g. 3GPP PUCCH 2).

(b) Low DMRS density.

Figure 3.2: General resource mapping: 1 OFDM symbol.

Furthermore, the transmitted signal $\mathbf{x}$ typically consists of data-dependent $\mathbf{x}^{(\mathrm{d})}$ and data-independent $\mathbf{x}^{(\mathrm{p})}$ components, known as pilot or reference signals. The reference signals are used to resolve channel ambiguity across time, frequency, and/or spatial domains. 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 $12KL$. This represents the number of complex dimensions or resource elements(REs) in the physical resource blocks (PRBs). The number of PRBs, K ranges from 1 to 16, while the number of OFDM symbols, L, ranges from 1 to 14, and can be increased if multiple slots are used for signaling the channel bits. Resource elements $\mathbf{x} = \{\mathrm{x}_n : n = 1, 2, \ldots, N\}$ are mapped onto N subcarriers such that $\forall n \in \mathcal{N}_p \cup \mathcal{N}_d$ with $\mathcal{N}_p$ the set of subcarriers for DMRS and $\mathcal{N}_d$ the set of subcarriers for data. The assumption in this

work is that the data-dependent components $\mathbf{x}^{(\mathsf{d})}$ are generated from a binary code whose output is interleaved and subsequently mapped onto 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. We denote the E interleaved coded bits as $e_k, k = 0, 1, \cdots, E-1$. Adjacent $\log_2 |\boldsymbol{\chi}|$ bit-tuples are used to select the modulated symbols in the symbol alphabet. Unless otherwise stated, we assume that Gray mapping is applied when using non-binary modulation schemes. In instance of the quadrature phase-shift keying (QPSK) modulation, pairs of bits $e[2n]$ and $e[2n+1]$, are mapped to complex-valued modulation symbols as follows:

$$\mathrm{x}^{(\mathsf{d})}[n] = \begin{cases} \frac{1}{\sqrt{2}}(1+j), & \text{if } (e[2n], e[2n+1]) = (0,0) \\ \frac{1}{\sqrt{2}}(-1+j), & \text{if } (e[2n], e[2n+1]) = (0,1) \\ \frac{1}{\sqrt{2}}(-1-j), & \text{if } (e[2n], e[2n+1]) = (1,1) \\ \frac{1}{\sqrt{2}}(1-j), & \text{if } (e[2n], e[2n+1]) = (1,0) \end{cases} \quad \forall n \in \mathcal{N}_d. \tag{3.1}$$

Likewise, it is worth examining how DMRS sequences making up $\mathbf{x}^{(\mathsf{p})}$ are constructed. These sequences are derived from Zadoff-Chu (ZC) sequences, leveraging their desirable properties in terms of ideal auto-correlation, low PAPR (Peak-to-Average Power Ratio), constant amplitude, and their proven standardization adoption in LTE and 5G NR. In that respect, the DMRS sequence is defined as:

$$\mathrm{x}^{(\mathsf{p})}[n] = \begin{cases} \exp\left(-j\pi \cdot \frac{un^2}{N_p}\right) & \text{if } N_p \text{ even} \\ \exp\left(-j\pi \cdot \frac{un(n+1)}{N_p}\right) & \text{if } N_p \text{ odd} \end{cases}, \forall n \in \mathcal{N}_p, \tag{3.2}$$

where u is the root index, which must be coprime with N_p (i.e., $\mathtt{gcd}(u, N_p) = 1$). The power of each DMRS sequence is typically normalized to unity.

3 BICM Receiver Design for 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:

$$\mathbf{y}_i = \mathbf{h}_i \mathbf{x} + \mathbf{z}_i, \quad i = 0, 1, \ldots, \mathsf{N_R} - 1, \tag{3.3}$$

where $\mathbf{y}_i$ represents an observed vector in N complex dimensions, $\mathbf{x}$ is an N-dimensional modulated vector transporting B channel bits, so that the messag,$m = 0, 1 \ldots, 2^B - 1$, $\mathbf{z}_i$ is additive white Gaussian noise whose real and imaginary components are independent and have variance σ^2 in each dimension.

3.1 Perfect Channel State Information

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

$$\mathrm{q}\left(\mathbf{x}, \{\mathbf{y}_i, \mathbf{h}_i\}\right) = \mathrm{p}\left(\mathbf{y}_i, \mathbf{h}_i \mid \mathbf{x}\right) = \mathrm{p}\left(\mathbf{y}_i \mid \mathbf{x}, \mathbf{h}_i\right) \mathrm{p}\left(\mathbf{h}_i \mid \mathbf{x}\right). \tag{3.4}$$

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 (3.4) can be dropped since it will disappear in (3.5). The likelihood function is commonly equivalent to

$$\mathrm{q}\left(\mathbf{x}, \{\mathbf{y}_i, \mathbf{h}_i\}\right) \propto \exp\left(-\frac{1}{\mathrm{N}_0} ||\mathbf{y}_i - \mathbf{h}_i \mathbf{x}||^2\right). \tag{3.5}$$

Using the norm extension property: $||\mathbf{y}_i - \mathbf{h}_i \mathbf{x}||^2 = ||\mathbf{y}_i||^2 + ||\mathbf{h}_i \mathbf{x}||^2 - 2\mathrm{Re}\left(\mathbf{y}_i \mathbf{h}_i^\dagger \mathbf{x}^\dagger\right)$, and ignoring terms that are independent of $\mathbf{x}$, the likelihood function then simply becomes:

$$\mathrm{q}\left(\mathbf{x}, \{\mathbf{y}_i, \mathbf{h}_i\}\right) \propto \quad \exp\left(\frac{2}{\mathrm{N}_0} \mathrm{Re}\left(\mathbf{y}_i \mathbf{h}_i^\dagger \mathbf{x}^\dagger\right) - \frac{1}{\mathrm{N}_0} ||\mathbf{h}_i \mathbf{x}||^2\right). \tag{3.6}$$

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

$$\mathrm{q}\left(e_j(\mathbf{x}) = b, \{\mathbf{y}_i, \mathbf{h}_i\}\right) = \sum_{\mathbf{x} \in \mathcal{X}_b^j} \mathrm{q}\left(\mathbf{x}, \{\mathbf{y}_i, \mathbf{h}_i\}\right). \tag{3.7}$$

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

$$\Lambda^j\left(\mathbf{y}_i\right) = \log \frac{\mathrm{q}\left(e_j(\mathbf{x}) = 0, \{\mathbf{y}_i, \mathbf{h}_i\}\right)}{\mathrm{q}\left(e_j(\mathbf{x}) = 1, \{\mathbf{y}_i, \mathbf{h}_i\}\right)}. \tag{3.8}$$

We simplify (3.8) using a *max-log approximation*: $\log\left\{\sum_i \exp\left(\lambda_i\right)\right\} \sim \max_i\left\{\lambda_i\right\}$, resulting in

$$\Lambda^j\left(\mathbf{y}\right) = \sum_{b \in \{0,1\}} (-1)^b \left[\max_{\mathbf{x} \in \mathcal{X}_b^j} \frac{1}{\mathrm{N}_0} \left(\sum_{i=0}^{\mathsf{N_R}-1} 2\,\mathrm{Re}\left(\mathbf{y}_r \mathbf{h}_i^\dagger \mathbf{x}^\dagger\right) - ||\mathbf{h}_i \mathbf{x}||^2\right)\right]. \tag{3.9}$$

This metric is typically used in *Perfect CSI* based receivers as well as in conventional quasi-coherent receivers which employs least-squares channel estimation followed by linear interpolation, thus by substituting $\mathbf{h}$, with $\hat{\mathbf{h}}$. We consider these receivers as a benchmark for comparison with the subsequent proposed receivers.

Moreover, within the framework of a conventional receiver, it is presupposed that, at the very least, the observation of a single reference signal spans a PRB in order to generate the coded bits corresponding to each data symbol in that PRB. In addition, this conventional metric applies a so-called *symbol-by-symbol* detection, that's to say that each symbol is detected independently of the other symbols. We will note this as *No CSI* Conv $(\cdots \otimes \cdots)$.

3.2 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* (NLOS) components. The overall unknown channel observation at the $i-th$ receiver antenna port is given by

$$\mathbf{h}_i = \left(\sqrt{\alpha} e^{j\theta_i} + \sqrt{1-\alpha}\mathrm{h}_i^{(f)}\right)\mathbf{I}, \quad \forall i, \left\{\left(\theta_i,\ \mathrm{h}_i^{(f)}\right)\right\}_{i=1}^{\mathsf{N_R}} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}[0, 2\pi) \times \mathbb{CN}(0,1), \tag{3.10}$$

where θ_i is assumed to be IID 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 therefore Ricean distributed.

Definition 3.1. (Non-coherent fading channels.)
Non-coherent fading channels refer to fading channels in which the fading coefficients are not known a priori and require estimation.

Definition 3.2. (Ricean distributed channel.)
Assuming that $|\mathrm{h}_i|$ follows a Rice distribution, we can define the parameters ($\texttt{Rician factor}: K, \texttt{mean}: \mu, \texttt{variance}: \sigma_h^2$) with respect to α. Any Rice distributed channel model can be represented as :

$$\mathbf{h}_i \triangleq \left(\sqrt{\frac{K}{1+K}}\, e^{j\theta_i} + \sqrt{\frac{1}{1+K}}\, \mathrm{h}_i^{(f)}\right)\mathbf{I}, \tag{3.11}$$

where $\mu \triangleq \sqrt{\frac{K}{2(1+K)}}, \quad \sigma_h^2 \triangleq \sqrt{\frac{1}{2(1+K)}}, \alpha = \frac{K}{1+K}, 1-\alpha = \frac{1}{1+K}$.

Definition 3.3. (Line-of-Sight channel.)
A LOS channel is a wireless communication link that connects a transmitter to a receiver without any barriers in between. This means there are no obstacles like buildings or hills blocking the signal's path, allowing it to travel directly from one point to another.

Furthermore, 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.

Proposition 3.1. (A novel soft-likelihood metric for SIMO.)

Neglecting multiplicative terms independent of the transmitted message, the likelihood function can be expressed as follows:

$$\mathrm{q}\left(\mathbf{x},\mathbf{y}\right)=\prod_{i=0}^{\mathsf{N_R}-1}\frac{1}{\mathbf{L}_{\mathsf{x}}}\exp\left(-\frac{\alpha\left\|\mathbf{x}\right\|^{2}}{\mathbf{L}_{\mathsf{x}}}+\beta_{x}\left|\mathbf{x}^{\dagger}\mathbf{y}_{i}\right|^{2}\right)\times\mathrm{I}_{0}\left(\frac{2\sqrt{\alpha}}{\mathbf{L}_{\mathsf{x}}}\left|\mathbf{x}^{\dagger}\mathbf{y}_{i}\right|\right),\tag{3.12}$$

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

Proof. Since θ_i is unknown and randomly distributed over $[0,2\pi)$, the conditional probability density function can be written as follows:

$$\mathrm{q}\left(\mathbf{x},\mathbf{y}_i\right)=\frac{1}{2\pi\det\Phi}\int_0^{2\pi}\exp\left(-\frac{1}{2}\left(\mathbf{y}_i-\ \mu\{\mathbf{x},\theta_i\}\right)^{\dagger}\Phi^{-1}\left(\mathbf{y}_i-\sqrt{\alpha}e^{j\theta_i}\mathbf{x}\{\mathbf{x},\theta_i\}\right)\right)\mathrm{d}\theta_i\ .\tag{3.13}$$

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

$$\mathrm{q}\left(\mathbf{x},\mathbf{y}_i\right)=\frac{1}{2\pi\det\Phi}\int_0^{2\pi}\exp\left(-\frac{1}{2}\left(\mathbf{y}_i-\sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)^{\dagger}\Phi^{-1}\left(\mathbf{y}_i-\sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)\right)\mathrm{d}\theta_i\ .\tag{3.14}$$

Covariance Matrix

Knowing that

$$\mathbf{y}_i-\sqrt{\alpha}e^{j\theta_i}\mathbf{x}=\sqrt{1-\alpha}\mathrm{h}_i^{(f)}\mathbf{x}+\mathbf{z}_i,\tag{3.15}$$

Then

$$\begin{aligned}\Phi&\triangleq\frac{1}{2}\mathbb{E}\left[\left(\sqrt{1-\alpha}\mathrm{h}_i^{(f)}\mathbf{x}+\mathbf{z}_i\right)\left(\sqrt{1-\alpha}\mathrm{h}_i^{(f)}\mathbf{x}+\mathbf{z}_i\right)^{\dagger}\right]\\&=\frac{1}{2}\mathbb{E}\left[\left(\sqrt{1-\alpha}\mathrm{h}_i^{(f)}\mathbf{x}+\mathbf{z}_i\right)\left(\sqrt{1-\alpha}\mathbf{x}^{\dagger}\mathrm{h}_i^{(f)\dagger}+\mathbf{z}_i^{\dagger}\right)\right]\\&=(1-\alpha)\frac{1}{2}\mathbb{E}\left[\mathrm{h}_i^{(f)}\mathbf{x}\mathrm{x}^{\dagger}\mathrm{h}_i^{(f)\dagger}\right]+\frac{1}{2}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i^{\dagger}\right]\\&=(1-\alpha)\mathbf{x}\mathbf{x}^{\dagger}\sigma_h^2+\sigma_z^2\mathbf{I},\text{where }\sigma_h^2=1\\&=(1-\alpha)\mathbf{x}\mathbf{x}^{\dagger}+\sigma_z^2\mathbf{I}.\end{aligned}\tag{3.16}$$

Determinant

$$\begin{aligned}\det\Phi&=\det\left((1-\alpha)\mathbf{x}\mathbf{x}^{\dagger}+\sigma_z^2\mathbf{I}\right)\\&=\det\left(\sigma_z^2\mathbf{I}+(1-\alpha)\mathbf{x}\mathbf{x}^{\dagger}\right).\end{aligned}\tag{3.17}$$

Property 3.1. (Useful property for calculating the determinant)

1. Calculation of the determinant of the type $\det\left(\sigma_z^2\mathbf{I}+k\mathbf{x}\mathbf{x}^{\dagger}\right)$, where k is a scalar defined by $k=(1-\alpha)$, $\mathbf{x}$ and $\mathbf{x}^{\dagger}$ are complex column vectors (and thus $\mathbf{x}^{\dagger}$ is the conjugate transpose of $\mathbf{x}$), $\mathbf{I}$ is the identity matrix (of dimension $N\times N$), and σ_z^2 is a scalar (the variance of the Gaussian noise).
2. When we have a matrix of the form: $\mathbf{A}=\mathbf{B}+\mathbf{u}\mathbf{v}^{\dagger}$, where $\mathbf{u}$ and $\mathbf{v}^{\dagger}$ are column vectors, we can apply the determinant formula for a matrix perturbed by a rank-1 term (i.e., the rank-1 matrix determinant formula), where $\mathbf{B}$ is an invertible square matrix of dimension $N\times N$, $\mathbf{u}$ and $\mathbf{v}^{\dagger}$ are complex column vectors of dimension $N\times 1$, and $\mathbf{u}\mathbf{v}^{\dagger}$ is thus a rank-1 matrix (which is true for the product of a column vector and its conjugate transpose).
   The determinant of $\mathbf{A}$ is given by:

$$\det(\mathbf{A})\triangleq\det(\mathbf{B})\cdot\left(1+\mathbf{v}^{\dagger}\mathbf{B}^{-1}\mathbf{u}\right).\tag{3.18}$$

   Indeed, the term $\mathbf{u}\mathbf{v}^{\dagger}$ introduces a rank-1 perturbation to the matrix $\mathbf{B}$, and the formula shows that the effect on the determinant is simply a multiplicative factor, also called a corrective factor, involving $\left(1+\mathbf{v}^{\dagger}\mathbf{B}^{-1}\mathbf{u}\right)$.
3. Special Case: diagonal matrix and rank-1 perturbation.
   If $\mathbf{B}=\sigma_z^2\mathbf{I}$ (a scalar diagonal matrix with $\sigma_z^2>0$) and $\mathbf{u}=\mathbf{x}$ and $\mathbf{v}^{\dagger}=\mathbf{x}^{\dagger}$, then the matrix becomes: $\mathbf{A}=\sigma^2\mathbf{I}+k\mathbf{x}\mathbf{x}^{\dagger}$, where k is a scalar.

The determinant is then given by :

$$\det(\mathbf{A}) = \sigma_z^{2N}\left(1 + \frac{k\mathbf{x}^\dagger\mathbf{x}}{\sigma^2}\right), \tag{3.19}$$

where $\mathbf{x}^\dagger\mathbf{x} = \|\mathbf{x}\|^2$ is the square of the magnitude of the vector $\mathbf{x}$.

Taking into account this property regarding the calculation of the determinant, (3.17) then becomes,

$$\begin{aligned}\det\Phi &= \det\left(\sigma_z^2\mathbf{I} + (1-\alpha)\mathbf{x}\mathbf{x}^\dagger\right),\\ &= \sigma_z^{2N}\left(1 + \frac{(1-\alpha)\mathbf{x}^\dagger\mathbf{x}}{\sigma_z^2}\right),\\ &= \sigma_z^{2N}\left(1 + \frac{(1-\alpha)\|\mathbf{x}\|^2}{\sigma_z^2}\right).\end{aligned} \tag{3.20}$$

Let's try to simplify the determinant expression by ignoring the multiplicative constant terms

$$\begin{aligned}\det\Phi &= \sigma_z^{2N}\left(1 + \frac{(1-\alpha)\|\mathbf{x}\|^2}{\sigma^2}\right),\\ &= \left(\frac{\mathrm{N}_0}{2}\right)^N\left(1 + \frac{(1-\alpha)\|\mathbf{x}\|^2}{\frac{\mathrm{N}_0}{2}}\right),\\ &= \left(\frac{\mathrm{N}_0}{2}\right)^{N-1}\left(\frac{\mathrm{N}_0}{2} + (1-\alpha)\|\mathbf{x}\|^2\right),\\ &= \frac{1}{2}\left(\frac{\mathrm{N}_0}{2}\right)^{N-1}\left(\mathrm{N}_0 + 2(1-\alpha)\|\mathbf{x}\|^2\right),\\ &= c\left(\mathrm{N}_0 + 2(1-\alpha)\|\mathbf{x}\|^2\right), \text{ where } c = \frac{1}{2}\left(\frac{\mathrm{N}_0}{2}\right)^{N-1} \text{ is a constante,}\\ &\propto \mathrm{N}_0 + 2(1-\alpha)\|\mathbf{x}\|^2.\end{aligned} \tag{3.21}$$

In what follows, we will pose

$$\mathbf{L}_{\mathrm{x}} = \mathrm{N}_0 + 2(1-\alpha)\|\mathbf{x}\|^2. \tag{3.22}$$

In place of $\det\Phi$, we will use $\mathbf{L}_{\mathrm{x}}$ because the multiplicative constant term c can be ignored.

Inverse of Φ

The matrix inversion *Lemma* or *the Woodbury matrix identity* [138] must be used.

$$\Phi^{-1} \triangleq (\mathbf{A} + \mathbf{UCV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{C}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}\mathbf{A}^{-1}. \tag{3.23}$$

Note that, here, we have a special case where $\mathbf{V}$, $\mathbf{U}$ are vectors, consequently $\mathrm{rank}\{\mathbf{x}\mathbf{x}^\dagger\} = 1$.
Let's say :

$$\begin{cases}\mathbf{A} = \sigma_z^2\mathbf{I},\\ \mathbf{C} = (1-\alpha)\mathbf{I},\\ \mathbf{U} = \mathbf{x},\\ \mathbf{V} = \mathbf{x}^\dagger\end{cases}. \tag{3.24}$$

$$
\begin{aligned}
\Phi^{-1} &\triangleq (\mathbf{A}+\mathbf{UCV})^{-1}, \\
&= (\sigma_z^2\mathbf{I})^{-1} - (\sigma_z^2\mathbf{I})^{-1}\mathbf{x}\left[((1-\alpha)\mathbf{I})^{-1} + \mathbf{x}^\dagger(\sigma_z^2\mathbf{I})^{-1}\mathbf{x}\right]^{-1}\mathbf{x}^\dagger(\sigma_z^2\mathbf{I})^{-1}, \\
&= (\sigma_z^2\mathbf{I})^{-1}\left(\mathbf{I} - \mathbf{x}\left[((1-\alpha)\mathbf{I})^{-1} + \mathbf{x}^\dagger(\sigma_z^2\mathbf{I})^{-1}\mathbf{x}\right]^{-1}\mathbf{x}^\dagger(\sigma_z^2\mathbf{I})^{-1}\right), \\
&= (\sigma_z^2\mathbf{I})^{-1}\left(\mathbf{I} - \mathbf{x}\left[((1-\alpha)\mathbf{I})^{-1}(\sigma_z^2\mathbf{I}) + \mathbf{x}^\dagger\mathbf{x}\underbrace{(\sigma_z^2\mathbf{I})^{-1}(\sigma_z^2\mathbf{I})}_{\mathbf{I}}\right]^{-1}\mathbf{x}^\dagger\right), \\
&= (\sigma_z^2\mathbf{I})^{-1}\left(\mathbf{I} - \mathbf{x}\left[(1-\alpha)^{-1}\mathbf{I}(\sigma_z^2\mathbf{I}) + \mathbf{x}^\dagger\mathbf{x}\mathbf{I}\right]^{-1}\mathbf{x}^\dagger\right), \\
&= (\sigma_z^2\mathbf{I})^{-1}\left(\mathbf{I} - \mathbf{x}\left[\frac{\sigma_z^2}{(1-\alpha)}\mathbf{I} + \mathbf{x}^\dagger\mathbf{x}\mathbf{I}\right]^{-1}\mathbf{x}^\dagger\right), \text{where } \sigma_z^2\mathbf{I} = \frac{\mathrm{N}_0}{2}\mathbf{I}, \\
&= \frac{2}{\mathrm{N}_0}\mathbf{I}\left(\mathbf{I} - \mathbf{x}\left[\frac{\mathrm{N}_0}{2(1-\alpha)}\mathbf{I} + \mathbf{x}^\dagger\mathbf{x}\mathbf{I}\right]^{-1}\mathbf{x}^\dagger\right), \\
&= \frac{2}{\mathrm{N}_0}\mathbf{I} - \frac{2}{\mathrm{N}_0}\mathbf{I}\,\mathbf{x}\left[\frac{\mathrm{N}_0 + 2(1-\alpha)\left\|\mathbf{x}\right\|^2}{2(1-\alpha)}\mathbf{I}\right]^{-1}\mathbf{x}^\dagger, \\
&= \frac{2}{\mathrm{N}_0}\mathbf{I} - \frac{2}{\mathrm{N}_0}\mathbf{x}\left(\frac{2(1-\alpha)}{\mathrm{N}_0 + 2(1-\alpha)\left\|\mathbf{x}\right\|^2}\right)\mathbf{x}^\dagger.
\end{aligned}
\tag{3.25}
$$

Let say $\beta_x = \frac{2(1-\alpha)}{\mathrm{N}_0(\mathrm{N}_0+2(1-\alpha)\|\mathbf{x}\|^2)}$, then

$$
\Phi^{-1} = \frac{2}{\mathrm{N}_0} - 2\mathbf{x}\beta_x\mathbf{x}^\dagger. \tag{3.26}
$$

Likelihood function

$$
\begin{aligned}
\mathrm{q}\left(\mathbf{x},\mathbf{y}_i\right) &= \frac{1}{2\pi\det\Phi}\exp\left(-\frac{1}{2}\left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)^\dagger\left(\frac{2}{\mathrm{N}_0} - 2\mathbf{x}\beta_x\mathbf{x}^\dagger\right)\left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)\right), \\
&= \frac{1}{2\pi\det\Phi}\exp\left(-\frac{1}{\mathrm{N}_0}\left\|\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right\|^2 + \beta_x\left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)^\dagger\mathbf{x}\mathbf{x}^\dagger\left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)\right), \\
&= \frac{1}{2\pi\det\Phi}\exp\left(-\frac{1}{\mathrm{N}_0}\left\|\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right\|^2 + \left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)^\dagger\mathbf{x}\beta_x\mathbf{x}^\dagger\left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)\right), \\
&= \frac{1}{2\pi\det\Phi}\exp\left(-\frac{1}{\mathrm{N}_0}\left\|\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right\|^2 + \left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)^\dagger(\mathbf{x}^\dagger)^\dagger\beta_x\mathbf{x}^\dagger\left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)\right), \\
&= \frac{1}{2\pi\det\Phi}\exp\left(-\frac{1}{\mathrm{N}_0}\left\|\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right\|^2 + \left(\mathbf{x}^\dagger\left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)\right)^\dagger\beta_x\left(\mathbf{x}^\dagger\left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)\right)\right), \\
&= \frac{1}{2\pi\det\Phi}\exp\left(-\frac{1}{\mathrm{N}_0}\left\|\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right\|^2 + \beta_x\left\|\mathbf{x}^\dagger\left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)\right\|^2\right), \\
&= \frac{1}{2\pi\det\Phi}\exp\left(-\frac{1}{\mathrm{N}_0}\left\|\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right\|^2 + \beta_x\left\|\left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)^\dagger\mathbf{x}\right\|^2\right), \\
&= \frac{1}{2\pi\det\Phi}\exp\left(-\frac{1}{\mathrm{N}_0}\left\|\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right\|^2 + \beta_x\left\|\left(\mathbf{y}_i^\dagger - \sqrt{\alpha}e^{-j\theta_i}\mathbf{x}^\dagger\right)\mathbf{x}\right\|^2\right).
\end{aligned}
\tag{3.27}
$$

Let's try to expend terms in the exponential :

$$\begin{aligned}
A &= \frac{1}{\mathrm{N}_0}\left\|\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right\|^2 \\
&= \frac{1}{\mathrm{N}_0}\left\|\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right\|^2, \\
&= \frac{1}{\mathrm{N}_0}\left(\|\mathbf{y}_i\|^2 + \left\|\sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right\|^2 - 2\,\mathrm{Re}\left(\mathbf{y}_i\sqrt{\alpha}\mathrm{e}^{-j\theta_i}\mathbf{x}\dagger\right)\right), \\
&= \frac{1}{\mathrm{N}_0}\left(\|\mathbf{y}_i\|^2 + \alpha\|\mathbf{x}\|^2 - 2\sqrt{\alpha}\,\mathrm{Re}\left(e^{-j\theta_i}\mathbf{y}_i\mathbf{x}^\dagger\right)\right), \\
&= \frac{1}{\mathrm{N}_0}\|\mathbf{y}_i\|^2 + \frac{\alpha}{\mathrm{N}_0}\|\mathbf{x}\|^2 - \frac{2\sqrt{\alpha}}{\mathrm{N}_0}\,\mathrm{Re}\left(\mathbf{y}_i^\dagger\mathbf{x}e^{j\theta_i}\right), \\
&= \frac{1}{\mathrm{N}_0}\|\mathbf{y}_i\|^2 + \frac{\alpha}{\mathrm{N}_0}\|\mathbf{x}\|^2 - \frac{2\sqrt{\alpha}}{\mathrm{N}_0}\,\mathrm{Re}\left(|\mathbf{y}_i^\dagger\mathbf{x}|e^{j\phi_i}e^{j\theta_i}\right), \\
&= \frac{1}{\mathrm{N}_0}\|\mathbf{y}_i\|^2 + \frac{\alpha}{\mathrm{N}_0}\|\mathbf{x}\|^2 - \frac{2\sqrt{\alpha}}{\mathrm{N}_0}|\mathbf{x}^\dagger\mathbf{y}_i|\cos\left(\phi_i + \theta_i\right).
\end{aligned} \tag{3.28}$$

knowing that $\mathbf{x}^\dagger\mathbf{y}_i = |\mathbf{x}^\dagger\mathbf{y}_i|e^{j\phi_i}$, where $\phi_i = \angle\mathbf{x}^\dagger, \mathbf{y}_i$,

$$\begin{aligned}
B &= \beta_x\left\|\left(\mathbf{y}_i^\dagger - \sqrt{\alpha}e^{-j\theta_i}\mathbf{x}^\dagger\right)\mathbf{x}\right\|^2 \\
&= \beta_x\|\mathbf{x}\|^2\left\|\mathbf{y}_i^\dagger - \sqrt{\alpha}\mathrm{e}^{-j\theta_i}\mathbf{x}^\dagger\right\|^2, \\
&= \beta_x\|\mathbf{x}\|^2\left(\left\|\mathbf{y}_i^\dagger\right\|^2 + \left\|\sqrt{\alpha}\mathrm{e}^{-j\theta_i}\mathbf{x}^\dagger\right\|^2 - 2\,\mathrm{Re}\left(\mathbf{y}_i^\dagger\sqrt{\alpha}(e^{-j\theta_i})^\dagger(\mathbf{x}^\dagger)^\dagger\right)\right) \\
&= \beta_x\|\mathbf{x}\|^2\left(\left\|\mathbf{y}_i^\dagger\right\|^2 + \alpha\left\|e^{-j\theta_i}\mathbf{x}^\dagger\right\|^2 - 2\sqrt{\alpha}\,\mathrm{Re}\left(\mathbf{y}_i^\dagger e^{j\theta_i}\mathbf{x}\right)\right), \\
&= \beta_x\left|\mathbf{x}\mathbf{y}_i^\dagger\right|^2 + \alpha\beta_x\|\mathbf{x}\|^2\|\mathbf{x}\|^2 - 2\sqrt{\alpha}\beta_x\|\mathbf{x}\|^2\,\mathrm{Re}\left(\mathbf{y}_i^\dagger\mathbf{x}e^{j\theta_i}\right), \\
&= \beta_x\left|\mathbf{x}\mathbf{y}_i^\dagger\right|^2 + \alpha\beta_x\|\mathbf{x}\|^2\|\mathbf{x}\|^2 - 2\sqrt{\alpha}\beta_x\|\mathbf{x}\|^2\,\mathrm{Re}\left(|\mathbf{y}_i^\dagger\mathbf{x}|e^{j\phi_i}e^{j\theta_i}\right), \\
&= \beta_x\left|\mathbf{x}\mathbf{y}_i^\dagger\right|^2 + \alpha\beta_x\|\mathbf{x}\|^2\|\mathbf{x}\|^2 - 2\sqrt{\alpha}\beta_x\|\mathbf{x}\|^2|\mathbf{x}^\dagger\mathbf{y}_i|\cos\left(\phi_i + \theta_i\right).
\end{aligned} \tag{3.29}$$

Thus (-A) + (B) is

$$\begin{aligned}
-A + B = &-\frac{1}{\mathrm{N}_0}\|\mathbf{y}_i\|^2 - \frac{\alpha\|\mathbf{x}\|^2}{\mathrm{N}_0} + \frac{2\sqrt{\alpha}}{\mathrm{N}_0}|\mathbf{y}_i\mathbf{x}^\dagger|\cos\left(\phi_i + \theta_i\right), \\
&+ \beta_x\left|\mathbf{x}\mathbf{y}_i^\dagger\right|^2 + \alpha\beta_x\|\mathbf{x}\|^2\|\mathbf{x}\|^2 - 2\sqrt{\alpha}\beta_x\|\mathbf{x}\|^2|\mathbf{x}^\dagger\mathbf{y}_i|\cos\left(\phi_i + \theta_i\right).
\end{aligned} \tag{3.30}$$

Ignoring the independent terms of $\mathbf{x}$, then combining the common terms, we get

$$\begin{aligned}
&-\frac{1}{\mathrm{N}_0}\left\|\mathbf{y}_i - \sqrt{\alpha}\mathrm{e}^{j\theta_i}\mathbf{x}\right\|^2 + \beta_x\left\|\left(\mathbf{y}_i^\dagger - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}^\dagger\right)\mathbf{x}\right\|^2 = -\alpha||\mathbf{x}||^2\left(\frac{1}{\mathrm{N}_0} - \beta_x||\mathbf{x}||^2\right) + \\
&\beta_x\left|\mathbf{x}^\dagger\mathbf{y}_i\right|^2 + 2\sqrt{\alpha}\left(\frac{\mathbf{1}}{\mathrm{N}_0} - \beta_\mathbf{x}||\mathbf{x}||^2\right)\left|\mathbf{x}^\dagger\mathbf{y}_i\right|\cos\left(\phi_i + \theta_i\right).
\end{aligned} \tag{3.31}$$

Since θ is unknown, the likelihood function is equivalent to

$$\begin{aligned}
\mathrm{q}\left(\mathbf{x}, \mathbf{y}_i\right) &= \frac{1}{2\pi\det\Phi}\int_0^{2\pi}\exp\left(-\frac{1}{2}\left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)^\dagger\Phi^{-1}\left(\mathbf{y}_i - \sqrt{\alpha}e^{j\theta_i}\mathbf{x}\right)\right)\mathrm{d}\theta_i \\
&= \frac{1}{2\pi\det\Phi}\exp\left(-\alpha\|\mathbf{x}\|^2\left(\frac{1}{\mathrm{N}_0} - \beta_x\|\mathbf{x}\|^2\right) + \beta_x\left|\mathbf{x}^\dagger\mathbf{y}_i\right|^2\right)\times \\
&\int_0^{2\pi}\exp\left(2\sqrt{\alpha}\left(\frac{1}{\mathrm{N}_0} - \beta_\mathbf{x}||\mathbf{x}||^2\right)\left|\mathbf{x}^\dagger\mathbf{y}_i\right|\cos\left(\phi_i + \theta_i\right)\right)\mathrm{d}\theta_i,
\end{aligned} \tag{3.32}$$

knowing that $\mathrm{I}_0(\mathrm{z}) \triangleq \frac{1}{\pi}\int_{\theta_i=0}^{\pi}\exp(zcos(\lambda))\mathrm{d}\lambda$, where $\mathrm{I}_0(\cdot)$ is the zero-th order modified Bessel function of the first kind.

$$\begin{aligned}\mathrm{q}\left(\mathbf{x},\mathbf{y}_i\right) &= \frac{1}{2\pi\det\Phi}\exp\left(-\alpha\left\|\mathbf{x}\right\|^2\left(\frac{1}{\mathrm{N}_0}-\beta_x\left\|\mathbf{x}\right\|^2\right)+\beta_x\left|\mathbf{x}^\dagger\mathbf{y}_i\right|^2\right)\times\\ &\quad 2\pi\times\mathrm{I}_0\left(2\sqrt{\alpha}\left(\frac{1}{\mathrm{N}_0}-\beta_x\left\|\mathbf{x}\right\|^2\right)\left|\mathbf{x}^\dagger\mathbf{y}_i\right|\right),\\ &= \frac{1}{\det\Phi}\exp\left(-\alpha\left\|\mathbf{x}\right\|^2\left(\frac{1}{\mathrm{N}_0}-\beta_x\left\|\mathbf{x}\right\|^2\right)+\beta_x\left|\mathbf{x}^\dagger\mathbf{y}_i\right|^2\right)\times\\ &\quad \mathrm{I}_0\left(2\sqrt{\alpha}\left(\frac{1}{\mathrm{N}_0}-\beta_x\left\|\mathbf{x}\right\|^2\right)\left|\mathbf{x}^\dagger\mathbf{y}_i\right|\right).\end{aligned} \tag{3.33}$$

Then after ignoring multiplicative term that are independent of $\mathbf{x}$, it comes

$$\begin{aligned}\mathrm{q}\left(\mathbf{x},\mathbf{y}_i\right) &= \frac{1}{\mathbf{L}_{\mathrm{x}}}\exp\left(-\alpha\left\|\mathbf{x}\right\|^2\left(\frac{1}{\mathrm{N}_0}-\beta_x\left\|\mathbf{x}\right\|^2\right)+\beta_x\left|\mathbf{x}^\dagger\mathbf{y}_i\right|^2\right)\times\\ &\quad \mathrm{I}_0\left(2\sqrt{\alpha}\left(\frac{1}{\mathrm{N}_0}-\beta_x\left\|\mathbf{x}\right\|^2\right)\left|\mathbf{x}^\dagger\mathbf{y}_i\right|\right).\end{aligned} \tag{3.34}$$

Knowing that $\mathbf{L}_{\mathrm{x}}=\mathrm{N}_0+2(1-\alpha)\left\|\mathbf{x}\right\|^2$, let's try to find a relationship between β_x and $\mathbf{L}_{\mathrm{x}}$ to simplify the expression of the likelihood function.
Expressing β_x w.r.t. $\mathbf{L}_{\mathrm{x}}$, we have:

$$\begin{aligned}\beta_x &= \frac{2(1-\alpha)}{\mathrm{N}_0(\mathrm{N}_0+2(1-\alpha)\left\|\mathbf{x}\right\|^2)}\\ &= \frac{2(1-\alpha)}{\mathrm{N}_0\mathbf{L}_{\mathrm{x}}},\\ &= \frac{\mathbf{L}_{\mathrm{x}}-\mathrm{N}_0}{\left\|\mathbf{x}\right\|^2\mathrm{N}_0\mathbf{L}_{\mathrm{x}}},\\ &= \frac{\mathbf{L}_{\mathrm{x}}}{\left\|\mathbf{x}\right\|^2\mathrm{N}_0\mathbf{L}_{\mathrm{x}}}-\frac{\mathrm{N}_0}{\left\|\mathbf{x}\right\|^2\mathrm{N}_0\mathbf{L}_{\mathrm{x}}},\\ &= \frac{1}{\left\|\mathbf{x}\right\|^2\mathrm{N}_0}-\frac{1}{\left\|\mathbf{x}\right\|^2\mathbf{L}_{\mathrm{x}}}.\end{aligned} \tag{3.35}$$

$$\begin{aligned}\mathrm{q}\left(\mathbf{x},\mathbf{y}_i\right) = \frac{1}{\mathbf{L}_{\mathrm{x}}}\exp\left(-\alpha\left\|\mathbf{x}\right\|^2\left(\frac{1}{\mathrm{N}_0}-(\frac{1}{\cancel{\left\|\mathbf{x}\right\|^2}\mathrm{N}_0}-\frac{1}{\cancel{\left\|\mathbf{x}\right\|^2}\mathbf{L}_{\mathrm{x}}})\cancel{\left\|\mathbf{x}\right\|^2}\right)+\beta_x\left|\mathbf{x}^\dagger\mathbf{y}_i\right|^2\right)\\ \mathrm{I}_0\left(2\sqrt{\alpha}\left(\frac{1}{\mathrm{N}_0}-(\frac{1}{\cancel{\left\|\mathbf{x}\right\|^2}\mathrm{N}_0}-\frac{1}{\cancel{\left\|\mathbf{x}\right\|^2}\mathbf{L}_{\mathrm{x}}})\cancel{\left\|\mathbf{x}\right\|^2}\right)\left|\mathbf{x}^\dagger\mathbf{y}_i\right|\right).\end{aligned} \tag{3.36}$$

Thus the likelihood function is simply:

$$\mathrm{q}\left(\mathbf{x},\mathbf{y}_i\right) = \frac{1}{\mathbf{L}_{\mathrm{x}}}\exp\left(-\frac{\alpha\left\|\mathbf{x}\right\|^2}{\mathbf{L}_{\mathrm{x}}}+\beta_x\left|\mathbf{x}^\dagger\mathbf{y}_i\right|^2\right)\times\mathrm{I}_0\left(\frac{2\sqrt{\alpha}}{\mathbf{L}_{\mathrm{x}}}\left|\mathbf{x}^\dagger\mathbf{y}_i\right|\right). \tag{3.37}$$

■

Then we succinctly use (3.7-3.8) 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.* [88], 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 3.1. Moreover, the LLR bit metric calculation in log 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 LLR for the $j-th$ coded bit is given by (3.38).

$$\Lambda^j(\mathbf{y}) = \sum_{b\in\{0,1\}} (-1)^b \left[\max_{\mathbf{x}\in\mathbf{x}_b^j} \Big(\sum_{i=0}^{\mathsf{N_R}-1} \psi(\mathbf{x},\mathbf{y}_i) \Big) - \sum_{\mathbf{x}\in\mathbf{x}_b^j} \mathsf{N_R} \log \mathbf{L}_{\mathsf{x}} \right], \tag{3.38}$$

with the kernel function :

$$\psi(\mathbf{x},\mathbf{y}_i) \triangleq -\frac{\alpha \left\|\mathbf{x}\right\|^2}{\mathbf{L}_{\mathsf{x}}} + \beta_x \left|\mathbf{x}^\dagger \mathbf{y}_i\right|^2 + \frac{2\sqrt{\alpha}}{\mathbf{L}_{\mathsf{x}}} \left|\mathbf{x}^\dagger \mathbf{y}_i\right|.$$

Remark 3.1. In equation (3.38), 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. Furthermore, in the presence of strong LOS channels, the quadratic terms in equation (3.38) can also be ignored. 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 antennas ($\mathsf{N_R}$), 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 increase in computational complexity.

3.2.1 Joint Estimation and Detection

Observing the structure of the metrics and the absence of overlap between the data and the DMRS symbols, we can easily see that the channel estimate is part of the metric.

By posing $\mathbf{x}$ as a composite signal x comprising both data and dmrs in an interleaved fashion within a common OFDM symbol, it comes :$\mathbf{x} = \mathbf{x}^{(\mathsf{p})} + \mathbf{x}^{(\mathsf{d})}$ where d and p are subscripts representing data, DMRS components, respectively, we can highlight $\hat{\mathbf{h}}^{\mathrm{LS}}$ through the metric expression:

$$\begin{aligned}
\left|\mathbf{x}^\dagger \mathbf{y}_i\right| &= \left| \left(\mathbf{x}^{(\mathsf{p})} + \mathbf{x}^{(\mathsf{d})}\right)^\dagger \left(\mathbf{y}^{(\mathsf{p})} + \mathbf{y}^{(\mathsf{d})}\right) \right|, \\
&= \left| \left(\mathbf{x}^{(\mathsf{p})\dagger} + \mathbf{x}^{(\mathsf{d})\dagger}\right) \left(\mathbf{y}^{(\mathsf{p})} + \mathbf{y}^{(\mathsf{d})}\right) \right|, \\
&= \left| \underbrace{\mathbf{x}^{(\mathsf{p})\dagger} \mathbf{y}_i^{(\mathsf{p})}}_{\text{channel estimate}} + \cancelto{0}{\mathbf{x}^{(\mathsf{p})\dagger} \mathbf{y}_i^{(\mathsf{d})}} + \cancelto{0}{\mathbf{x}^{(\mathsf{d})\dagger} \mathbf{y}_i^{(\mathsf{p})}} + \mathbf{x}^{(\mathsf{d})\dagger} \mathbf{y}_i^{(\mathsf{d})} \right|, \\
&= \left| \underbrace{\mathbf{x}^{(\mathsf{p})\dagger} \mathbf{y}_i^{(\mathsf{p})}}_{\text{channel estimate}} + \mathbf{x}^{(\mathsf{d})\dagger} \mathbf{y}_i^{(\mathsf{d})} \right|.
\end{aligned} \tag{3.39}$$

Formally, we can therefore deduce the following relationship:

$$\begin{aligned}
\mathbf{x}^{(\mathsf{p})\dagger} \mathbf{y}_i^{(\mathsf{p})} &= \left(\mathbf{x}^{(\mathsf{p})\dagger} \mathbf{x}^{(\mathsf{p})}\right) \hat{\mathrm{h}}_i^{\mathrm{LS}}, \\
&= \|\mathbf{x}^{(\mathsf{p})}\|^2 \hat{\mathrm{h}}_i^{\mathrm{LS}} \\
&= N_p \rho \hat{\mathrm{h}}_i^{\mathrm{LS}},
\end{aligned} \tag{3.40}$$

where N_p is the number of pilots and ρ is the reference signal power and is typically normalized to unity.
Thus,

$$\left|\mathbf{x}^\dagger \mathbf{y}_i\right| = \left| N_p \hat{\mathrm{h}}_i^{\mathrm{LS}} + \mathbf{x}^{(\mathsf{d})\dagger} \mathbf{y}_i^{(\mathsf{d})} \right| . \tag{3.41}$$

The channel estimate $\hat{\mathrm{h}}_i^{\mathrm{LS}}$ is obtained via a joint *least-squares* (LS) channel estimation using averaging or smoothing over the number of dimensions exhibiting channel coherence (namely here, the coherence bandwidth).

Definition 3.4. (Channel coherence.)
Channel coherence is linked to two concepts:

- **Coherence time**: a time interval over which the channel gain or phase shift remains stable.
- **Coherence bandwidth**: a range of frequencies over which the channel remains constant (i.e., the amplitude and phase responses are constant).

In general, the channel estimation procedure will work as usual, and the resulting channel estimate is fed into the newly derived metrics considered here.

Proposition 3.2. (Joint estimation-detection metric)

Consequently, the derived likelihood function in (3.12) can be expressed in a way that enables joint estimation and detection, as follows:

$$\mathrm{q}\left(\mathbf{x},\mathbf{y}\right)=\prod_{i=0}^{\mathsf{N_R}-1}\frac{1}{\mathbf{L}_{\times}}\exp\left(-\frac{\alpha\left\|\mathbf{x}\right\|^{2}}{\mathbf{L}_{\times}}+\beta_{x}\left|N_{p}\hat{\mathrm{h}}_{i}^{\mathrm{LS}}+\ \mathbf{x}^{(\mathsf{d})\dagger}\mathbf{y}_{i}^{(\mathsf{d})}\right|^{2}\right)\times\mathrm{I}_{0}\left(\frac{2\sqrt{\alpha}}{\mathbf{L}_{\times}}\left|N_{p}\hat{\mathrm{h}}_{i}^{\mathrm{LS}}+\ \mathbf{x}^{(\mathsf{d})\dagger}\mathbf{y}_{i}^{(\mathsf{d})}\right|\right),\tag{3.42}$$

Corollary 3.2. In addition to proposition 3.2, an iterative joint estimation-detection can be considered for more reliable output LLRs. Indeed, the proposed metric in (3.39) can be extended to include iterative steps to enhance transmission quality. Here, we focus specifically on the estimation-detection aspect of the receiver. In general, the choice of where to apply the iterative process depends on the specific system requirements, complexity, available resources, transmission channel conditions, 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 SDT, 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.

The iterative process of the JED receiver comprises two successive stages. Firstly, the initial iterative process pertains to channel estimation, continuing until the maximum required number of iterations is reached. Subsequently, the second iterative process, related to the detector, follows until the maximum required number of iterations is attained. It is worth noting that the channel estimate involved in LLR generation corresponds to the channel estimate resulting from the last iteration or the maximum iteration in the joint channel estimation phase. This process can be viewed as a *2-layer iterative JED process*, with the first layer being iterative joint channel estimation and the second layer pertaining to iterative joint detection.

3.2.2 Block-wise Joint Estimation and Detection

For the case of polar or LDPC coded data, there is a convincing motivation to divide the coded streams into smaller blocks for detection for some complexity reasons. Assuming an ideal interleaving scenario with known channels [88], detection can be performed on individual modulated symbols. However, in the presence of joint estimation and detection, where interleaved DMRS and data symbols are considered, we need to deal with short blocks that contain both data or modulated symbols and DMRS symbols.

Hence, the received and transmitted signal sets are subdivided into smaller block segments, forming detection windows. Phrased directly, the newly proposed metric is applied to each group of symbols, or the so-called detection window. Noting that, in this proposed detection approach, we do not restrict the observation interval of a particular coded bit to the symbol in which it is conveyed.

Proposition 3.3. (Block-wise joint estimation - detection.)

The proposal consists of applying an advanced joint estimation-detection per block of $M=4$ modulated symbols forming a detection window to generate bit-level LLRs. Unlike conventional symbol-by-symbol detection in BICM systems where the observation for a particular coded bit is limited to the symbol in which it is conveyed, the proposed block-wise detection approach fundamentally leverages the statistical dependencies between adjacent symbols. In other respect, the LLR for a given coded bit incorporates information from all symbols within the detection window, rather than being constrained to its host symbol alone, thereby improving detection reliability.

To illustrate this statement, let $\mathbf{x}^{(\mathsf{d})}=\{\mathrm{x}_{n}^{(\mathsf{d})}\}_{n=1}^{M}\in\boldsymbol{\chi}^{M}$ be the candidate symbol vector corresponding to a block of M symbols (with $M=4$ in our case). Let the received frame be denoted by $\mathbf{Y}\in\mathbb{C}^{N\times\mathsf{N_R}}$, which contains both data and reference signals. We define the data-carrying positions (d) are divided into B non-overlapping blocks of size M, such that: $\mathsf{d}=\biguplus_{b=1}^{B}\mathcal{D}_{b},\quad$ with $|\mathcal{D}_{b}|=M$. For each block $b\in\{1,\ldots,B\}$, we extract the corresponding *block data observation matrix* from the received signal: $\mathbf{Y}_{b}^{(\mathsf{d})}\in\mathbb{C}^{M\times\mathsf{N_R}}$, which gathers all received data symbols at the M subcarriers in $\mathcal{D}_{b}$ across all $\mathsf{N_R}$ antennas. Each block $\mathbf{Y}_{b}^{(\mathsf{d})}$ is processed independently to compute the bit-wise likelihoods over the candidate symbol vectors. More precisely, unlike symbol-wise detection schemes, the proposed block-wise detection computes the LLRs of coded bits by evaluating joint symbol hypotheses over a detection window of size M. This involves assessing all possible candidate vectors $\mathbf{x}^{(\mathsf{d})}\in\boldsymbol{\chi}^{M}$, and partitioning them into two subsets ((i.e., grouping vectors where the bit is 0 vs. 1) based on the value of the j-th coded bit. Consequently, each LLR $\Lambda^{j}\left(\cdot\right)$ is derived by marginalizing the block-wise likelihood function.

Considering QPSK modulation and detection windows of $M = 4$, each jointly detected symbol block has $|\boldsymbol{\chi}|^M = 256$ possible candidate symbol vectors in the search space, where each candidate $\mathbf{x}^{(\mathsf{d})} \in \boldsymbol{\chi}^M$ represents a distinct symbol vector. In other terms, the detector exhaustively searches over all 256 possible symbol vector hypotheses to compute the LLRs.
The proposal is conceptually detailed in Algorithm 1.

Algorithm 2: Block-wise Joint Estimation-Detection (SIMO).

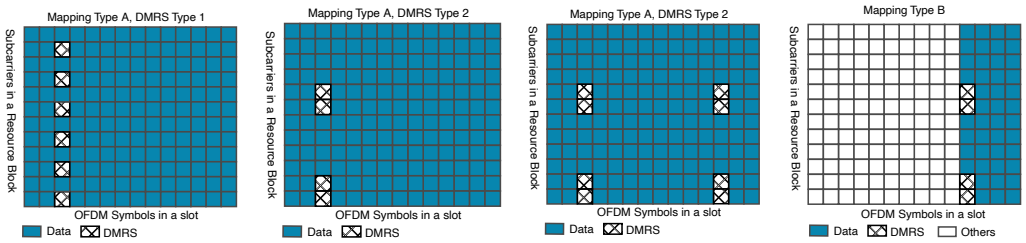
Input: Received frame: $\mathbf{Y} \in \mathbb{C}^{N \times \mathsf{N_R}}$, Data and DMRS indices: $(\mathsf{d}), (\mathsf{p})$, Candidate list size: $L = |\boldsymbol{\chi}|^M$, Candidate symbol set tensor: $\mathbf{X}^{(\mathsf{d})} \in \boldsymbol{\chi}^{M \times L}$, Adjustable parameter: α, noise power: N_0
Output: LLR vector $\boldsymbol{\Lambda}(\mathbf{y}^{(\mathsf{d})})$
for *each block* $b = 1, \ldots, B$ **do**
  Extract data observations over blocks:
  Let $\mathsf{idx}_b = [(b-1)\,M + 1 : b\,M]$ then,
  Set: $\mathbf{Y}_b^{(\mathsf{d})} = \mathbf{Y}^{(\mathsf{d})}\big(\mathsf{idx}_b, :\,\big) \in \mathbb{C}^{M \times \mathsf{N_R}}$.
  Set: $\mathbf{y}_{b,r}^{(\mathsf{d})} \coloneqq \mathbf{Y}_b^{(\mathsf{d})}(:, r) \in \mathbb{C}^{M \times 1}$.
  Step 1: Joint Channel Estimation
  for *each antenna* $r = 1, \ldots, \mathsf{N_R}$ **do**
    Extract DMRS observations:
    $\mathbf{y}_r^{(\mathsf{p})} = \mathbf{Y}^{(\mathsf{p})}(:, r) \in \mathbb{C}^{N_p \times 1}$
    Compute the direct channel estimate:
    $N_p \hat{\mathrm{h}}_r^{\mathrm{LS}} = \mathbf{x}^{(\mathsf{p})\dagger} \mathbf{y}_r^{(\mathsf{p})}$
  Step 2: Likelihood Function Processing
  for *each candidate index* $c = 1, \ldots, L$ **do**
    Set candidate symbol vector: $\mathbf{x}_c^{(\mathsf{d})} = \mathbf{X}^{(\mathsf{d})}(:, c)$
    Compute aggregated received signal energy:
    $\mathcal{S}_c = \sum_{r=1}^{\mathsf{N_R}} \left| N_p \hat{\mathrm{h}}_r^{\mathrm{LS}} + \mathbf{x}_c^{(\mathsf{d})\dagger} \mathbf{y}_{b,r}^{(\mathsf{d})} \right|^2$
    Set: $\mathbf{L}_{x_c} = \mathrm{N}_0 + 2(1-\alpha)\|\mathbf{x}_c^{(\mathsf{d})}\|^2$,
    Set: $\beta_{x_c} = \frac{2(1-\alpha)}{\mathrm{N}_0 \cdot \mathbf{L}_{x_c}}$
    Evaluate the likelihood function: $\mathrm{q}(\mathbf{x}_c^{(\mathsf{d})}, \mathbf{y}_b^{(\mathsf{d})}) = \frac{1}{\mathbf{L}_{x_c}} \exp\left(-\frac{\alpha \|\mathbf{x}_c^{(\mathsf{d})}\|^2}{\mathbf{L}_{x_c}} + \beta_{x_c} \cdot \mathcal{S}_c \right) \mathrm{I}_0\left(\frac{2\sqrt{\alpha \cdot \mathcal{S}_c}}{\mathbf{L}_{x_c}} \right)$
  Step 3: Joint Detection
  for *each bit index* $j = 1, \ldots, M \cdot \log_2 |\boldsymbol{\chi}|$ *in the block* **do**
    Compute the LLR of coded bit $e_j \in \{0, 1\}$:
    $\Lambda_j(\mathbf{y}_b^{(\mathsf{d})}) = \log \frac{\sum_{\mathbf{x}_c^{(\mathsf{d})} \in \boldsymbol{\chi}_0^j} \mathrm{q}(\mathbf{x}_c^{(\mathsf{d})}, \mathbf{y}_b^{(\mathsf{d})})}{\sum_{\mathbf{x}_c^{(\mathsf{d})} \in \boldsymbol{\chi}_1^j} \mathrm{q}(\mathbf{x}_c^{(\mathsf{d})}, \mathbf{y}_b^{(\mathsf{d})})}$
Step 4: Concatenate the block-wise LLR vectors
return $\boldsymbol{\Lambda}(\mathbf{y}^{(\mathsf{d})}) = \left[\Lambda(\mathbf{y}_1^{(\mathsf{d})}), \Lambda(\mathbf{y}_2^{(\mathsf{d})}), \ldots, \Lambda(\mathbf{y}_B^{(\mathsf{d})}) \right]$

In other respects, the computational complexity of the block-wise JED approach is typically on the order $\mathcal{O}\left(\mathsf{N_R} \times B \cdot |\boldsymbol{\chi}|^M \times M \cdot \log_2 |\boldsymbol{\chi}|\right)$.

3.2.3 JED Receiver on various PxSCH-based resources mapping setups

The proposal consists of an extension of the JED oriented receiver for different mapping types compliant with 5G NR resource grids. As an illustration, we use the PUSCH formats with DMRS mapping Type A, DMRS Type 1 and DMRS Type 2 mapping, where the DMRS are embedded at least in a single OFDM symbol, traditionally $\ell = 2, 3$, with the other OFDM symbols embedding the data or modulated symbols.
Furthermore, in 5G NR Front loaded DMRS technique aims for faster and efficient demodulation (reduce latency). Type 1 is much denser in frequency domain, as DMRS occupies is 50% of the RE's. In this case, the detection process envisaged is such as :

1. we will use the JED principle illustrated in proposition 3.2, to detect the data in the OFDM symbol(s) consisting of frequency interleaving between the data symbol and the DMRS symbol,
2. for other OFDM symbols, in particular those containing only data elements, the conventional pilot-assisted transmission detection method can be used.



(a) First DM-RS in symbol 2 or 3 of a slot.

(b) First DM-RS in symbol 2 or 3 of a slot.

(c) First DM-RS in symbol 2 and 11 a slot.

(d) First DM-RS in first symbol in the allocation PUSCH partially mapped to a slot.

Figure 3.3: DMRS PUSCH based resource mapping for JED oriented receiver.

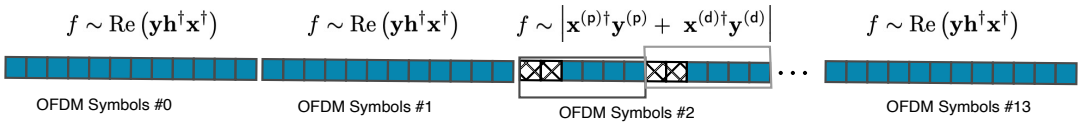


Figure 3.4: Conceptual Illustration of the JED principle across multiple OFDM symbols in a slot, f represents a dependent function relationship.

Proposition 3.4. (JED Receiver METRIC on PxSCH-type transmission)
We can thus summarise the proposition by referring to two different likelihood functions depending on the OFDM symbol processed by the demodulator in SISO BICM OFDM, such as :
Let's pose : $\mathcal{S}_{i,\ell} \triangleq \mathbf{x}_{\ell}^{(\mathsf{p})\dagger}\mathbf{y}_{i,\ell}^{(\mathsf{p})} + \mathbf{x}_{\ell}^{(\mathsf{d})\dagger}\mathbf{y}_{i,\ell}^{(\mathsf{d})}$

$$\mathrm{q}\left(\mathbf{x}_{\ell},\mathbf{y}_{\ell}\right) = \begin{cases} \prod_{i=0}^{\mathrm{N_R}-1} \frac{1}{\mathbf{L}_{\mathsf{x}}} \exp\left(-\frac{\alpha\left\|\mathbf{x}_{\ell}\right\|^2}{\mathbf{L}_{\mathsf{x}}} + \beta_x\left|\mathcal{S}_{i,\ell}\right|^2\right) \times \mathrm{I}_0\left(\frac{2\sqrt{\alpha}}{\mathbf{L}_{\mathsf{x}}}\left|\mathcal{S}_{i,\ell}\right|\right), & \texttt{if} \quad \ell = 2, \\ \prod_{i=0}^{\mathrm{N_R}-1} \exp\left(\frac{2}{\mathrm{N}_0}\mathrm{Re}\left(\mathbf{y}_{i,\ell}\hat{\mathbf{h}}_i^{\dagger}\mathbf{x}_{\ell}^{\dagger}\right) - \frac{1}{\mathrm{N}_0}||\hat{\mathbf{h}}_i\mathbf{x}_{\ell}||^2\right), & \texttt{if} \quad \ell = 0,1,3,\ldots,13, \end{cases} \tag{3.43}$$

where $\hat{\mathbf{h}}_i \triangleq \frac{1}{\mathcal{L}'}\sum_{\ell=1}^{\mathcal{L}'}(\mathbf{x}_{\ell}^{(\mathsf{p})\dagger}\mathbf{x}_{\ell}^{(\mathsf{p})})^{-1}x_{\ell}^{(\mathsf{p})\dagger}\mathbf{y}_{i,\ell}^{(\mathsf{p})}$.
As it stands, such a receiver may be viewed as running a detection diversity, including the advanced JED receiver and pilot-assisted conventional receiver, making it a diversity-based detection oriented receiver.
Let's now illustrate how to manage the estimation detection through the OFDM symbols, as illustrated in Figure 3.3b. The channel estimate is obtained as in the conventional method via least square either by averaging over the bandwidth of the signal assuming a flat fading or LOS type channel, or by interpolation (nearest, linear, or spline cubic) to better track the channel in the case where its frequency selectivity is at concern.
Furthermore, if among the OFDM symbols making up the same slot, some (i.e., $\mathcal{L}' \leq \mathcal{L}$ symbols) have the same pilot model as shown in Figure 3.3d, we can also average them in the time domain, in the channel estimation process, then $\mathcal{L}'$ is simply the average coefficient in the time domain.

3.2.4 Finit-Blocklength Bounds

Herein, we present the finite-blocklength information theory tools. An outer bound, derived from the metaconverse theorem (cf. [127, Th. 28]), is introduced, while an inner bound is established using the RCUs bound [130, Th. 1].

Theorem 3.1. [130, RCU bound, Th. 1].
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

$$\imath_s\left(\mathbf{x},\mathbf{y}\right) \triangleq \ln\frac{\mathrm{q}\left(\mathbf{x},\mathbf{y}\right)^s}{\mathbb{E}\left[\mathrm{q}\left(\mathbf{X}',\mathbf{y}\right)^s\right]}. \tag{3.44}$$

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:

$$\varepsilon \leq \inf_{s>0} \mathbb{E}\left[e^{-\left[\imath_s(\mathbf{x},\mathbf{y})-\ln\left(2^{R\times N}-1\right)\right]^+}\right], \tag{3.45}$$

where $[u]^+ \triangleq \max(0,u)$.
The maximum likelihood decoding metric is shown to be $\mathrm{q}(\mathbf{x},\mathbf{y}) = p_{\mathbf{Y}|\mathbf{X},\mathbf{H}}(\mathbf{y} \mid \mathbf{x},\mathbf{h})$. Thus, the underlying decoding metric is expressed as $\mathrm{q}(\mathbf{x},\mathbf{y}) \propto \exp\left(-\frac{1}{\mathrm{N}_0}||\mathbf{y}-\mathbf{h}\mathbf{x}||^2\right)$.

Theorem 3.2. [127, Metaconverse (Verdú–Han) bound, Th. 28].
To set the converse bound, consider :

$$\jmath_s(\mathbf{x},\mathbf{y}) \triangleq \ln \frac{\mathrm{q}(\mathbf{x},\mathbf{y})^s}{\mathbb{E}\left[\mathrm{q}(\mathbf{X}',\mathbf{y})^s\right]^{1/s}} = \frac{1}{s}\imath_s(\mathbf{x},\mathbf{y}) \tag{3.46}$$

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

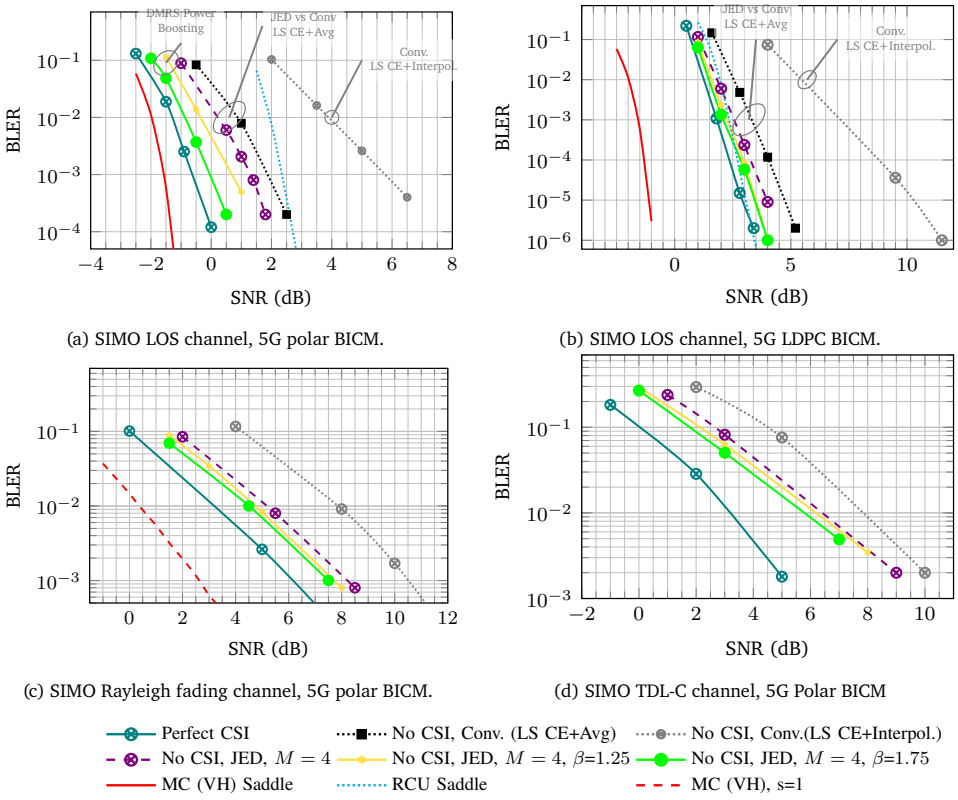
$$\varepsilon \geq \sup_{s>0} \max_{\lambda} \mathrm{P}\left[\jmath_s(\mathbf{x},\mathbf{y}) \leq \lambda\right] - e^{\lambda - R\times N}. \tag{3.47}$$

4 Numerical Results

For illustrative purposes, we consider two distinct configurations: (1×4) SIMO BICM, and (4×4) MIMO BICM. The subsequent figures illustrate the performance of BIPCM/BILCM with joint estimation and detection using the proposed block-wise joint estimation–detection metrics under various channel conditions. The evaluation highlights performance differences across three scenarios: *Perfect CSI* or ideal or coherent receiver, *No CSI* JED using the block-wise($M=4$) joint detection approach, and *No CSI* Conv, using conventional symbol-by-symbol detection.
The simulations employ NR POLAR and NR LDPC coding schemes paired with QPSK modulation. The transmission involves a transport block of 48 bits. The resource allocation procedure uses a single OFDM symbol with 48 resource elements spread over 4 PRBs (comprising 32 REs for data component and fewer REs for DMRS component).

4.1 Performance Analysis

The results illustrated in Figure 3.5a Figure 3.5c delineate the performance of the (1×4) SIMO BICM systems, with joint estimation and detection, over LOS channel ($\alpha=1$) and rayleigh fading channel ($\alpha=0$), respectively. This evaluation is aiming to discern the performance disparity between the *Perfect CSI* and *No CSI* conditions, particularly in coverage scenarios characterized by low signal-to-noise ratios. Moreover, we explore the most challenging situation where the DMRS density is very low, in order to show how robust the proposed receiver metrics can be compared to the conventional ones. To do this, we consider a single DMRS symbol across the entire coherence bandwidth or, at the very least, across a PRB. In this way, two structures of the conventional receiver are considered: (*No CSI Conv., LS CE + Interpol.*), where least squares channel estimation is followed by linear interpolation, and (*No CSI Conv., LS CE + Avg.*), where channel estimation is performed by averaging over the DMRS symbols for a fairer comparison with the proposed JED receiver.
Figure 3.5a illustrates the performance over a pure LOS channel ($\alpha=1$) under 5G polar coded modulation. The JED-based receiver achieves a performance gain of 3.75 dB and 0.75 dB compared to the conventional receivers at a BLER threshold of 1%. However, a gap of approximately 0.5 dB remains between the JED receiver (with $M=4$) and the perfect CSI receiver.
Similarly, Figure 3.5b shows the performance under 5G LDPC coded modulation. The JED-based receiver provides a performance gain of 3.5 dB over the conventional receiver (*No CSI Conv., LS CE + Interpolation*) at the same BLER threshold.
To further highlight the advantage of the JED receiver over the conventional receiver, Figure 3.5b depicts performance over a pure Rayleigh fading channel ($\alpha=0$) under 5G polar coded modulation. Here, the JED-based receiver ($M=4$) achieves a gain of 2.75 dB with respect to the conventional receiver (*No CSI Conv., LS CE + Interpolation*).
In addition, the Figures also includes finite blocklength bounds, integrating both converse and achievability results for a comprehensive comparative analysis. References are made to the scholarly works [127, 130] detailing the *metaconverse* (MC) and *Random Coding Union* (RCU) bounds employed. Indeed, The RCU provides an achievability bound and gives a practical estimate of how well a communication system can perform at finite blocklength. It



(a) SIMO LOS channel, 5G polar BICM.

(b) SIMO LOS channel, 5G LDPC BICM.

(c) SIMO Rayleigh fading channel, 5G polar BICM.

(d) SIMO TDL-C channel, 5G Polar BICM

Figure 3.5: 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, 4 DMRS), (1×4) SIMO, vs outer (MC) and inner (RCUs) bounds, DMRS power Boosting via a scaling factor β.

estimates the error probability of a random code at a given blocklength, rate, and SNR. Conversely, the meta-converse bound establishes a lower bound on the minimum achievable error probability for any coding scheme. In other words, it complements the finite blocklength achievability bound by setting a benchmark for the best possible performance, ensuring that practical systems do not overestimate their capabilities, which justifies its inclusion in the graphs.

As shown in Figure 3.5a, the Gallager parameter s is numerically optimized (i.e., $s \to s^*$) to maximize the RCU bound and minimize the meta-converse Verdú-Han bound (i.e., the tightest bound), where s^* corresponds to the saddle point. Conversely, in Figure 3.5c, no optimization of the Gallager parameter s is performed; instead, a fixed value of $s = 1$ is chosen, resulting in a more relaxed constraint.

Next, Figure 3.5d compares the performances under a TDL-C Non-Line-of-Sight (Long Delay Spread = 300 ns, urban macro, Sampling rate f_s = 30.72 MHz) channel configuration with 4 receive antennas. The antenna ports were subjected to independent and identically distributed realizations, with no correlation modeling applied. Indeed, the TDL-C, is a 3GPP reference channel model characterized by its long delay-spread and emphasis for non-MIMO evaluation scenarios [135]. It models a multipath-rich environment where no dominant direct path (LOS) is explicitly present. Specifically, the TDL channel, being purely NLOS, behaves like a multi-tap Rayleigh channel where all taps are statistically distributed as independent and random Rayleigh fading. Consequently, we implement the proposed metric by setting $\alpha = 0$ in our simulations. Since the TDL-C channel is frequency selective, channel estimation using the least squares method followed by interpolation is typically required to better track channel fluctuations. However, it is important to recall that for the JED receiver under consideration, no interpolation is performed. Instead, a least-squares (LS) estimation is conducted, followed by averaging over the dimension of DMRS symbols, before incorporating the resulting channel estimate into the soft detection metric.

Furthermore, Figure 3.6 displays LLR distribution characteristics shown via histograms at SNR = 0 dB for the *No CSI* Conv., the proposed *No CSI* JED ($M = 4$), and the *Perfect CSI* receivers. It should be noted that LLR values represent the confidence level in binary decisions (bit = 0 or 1) after demodulation. An LLR value ≈ 0 indicates significant uncertainty or ambiguity in the binary decision, whereas values $\gg 0$ or $\ll 0$ correspond to high-confidence decisions (bit = 1 or 0, respectively).

The histogram shapes thus reflect the quality of information provided by each receiver. The No CSI Conv receiver shows a histogram concentrated around zero, indicating weak LLR values. This demonstrates the receiver's low confidence in detected bits and consequently poor detection performance, which is characteristic of receivers operating without channel knowledge. In contrast, the No CSI JED ($M = 4$) receiver exhibits a bimodal histogram with peaks near -3 and $+3$, showing well-separated LLR values that correspond to more confident and accurate decisions. This strongly suggests effective receiver operation without explicit CSI, further supported by the close similarity between its histogram and that of the ideal receiver.

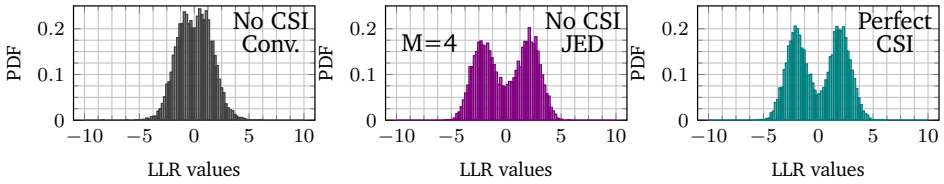


Figure 3.6: LLR distribution characteristics via histograms at SNR = 0 dB.

The discrepancy between the ideal receiver and the proposed one can be explained by the fact that the simulations are conducted under low DMRS density scenarios (i.e., 1 DMRS per PRB). In other respect, only fewer resource elements are allocated to DMRSs. This helps to some extent reduce the additional transmission overhead inherent in dense DMRS-assisted schemes.

Lastly, To fully exploit the performance potential of the JED receiver, we explore extra enhancement strategies. There are two possibilities: (a) DMRS densification, and (b) DMRS power boosting. If a certain sweet pot in terms of DMRS density per PRB cannot be found, increasing DMRS density inherently introduces additional transmission overhead. Therefore, the ideal approach is to prioritize transmission with low DMRS density just to bootstrap the JED receiver, followed by DMRS power boosting.

Conceptually, envision the signal as comprising a data component and a data-independent component, or pilots, in a frequency-interleaved fashion and both signals are jointly conveyed in common OFDM symbols. 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

$$\mathbf{x}_{\texttt{boosted}} = \mathbf{x}^{(\texttt{d})} + \beta \mathbf{x}^{(\texttt{p})}, \quad \text{s.t. } \beta = (1 + \zeta) \quad \forall \zeta \in (0, 1]. \tag{3.48}$$

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. Consequently, optimal performance enhancement is achieved when β is set to 1.75 (corresponding to a 75% increase in DMRS power w.r.t. its initial value). For instance, Figure 3.5a shows that the proposed JED-based receiver delivers an additional gain of approximately 1 dB, approaching the performance bound of the ideal receiver. Overall, 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.

Therefore, it can be also remarkably asserted that the advanced JED-based receiver outperforms the conventional counterpart and demonstrates greater resilience under imperfect channel estimation.

4.2 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 pinpoint the block size range wherein complexity is relatively low compared to conventional metrics in order to establish a better trade-off between performance and complexity.

Analytically, within a very short block regime(.i.e, small input symbol size), the conventional and JED receiver metrics will almost close. Considering the JED Metrics proposed in (3.12), we observe a quasi-linear complexity as a function of the input symbol size applied at the demodulator level, namely here $n = \{12, 24, 32, 48, 100\}$. Based on the average

time complexity curves on Figure 3.7, the complexity of the JED($M = 4$) receiver is of the form

$$\forall \varepsilon \in (0,1), \exists\, a \in \mathbb{R}_{>0} \quad \text{s.t. } \mathsf{T}(n) = n^{1+\varepsilon} = \mathcal{O}\left(n^{1+\varepsilon}\right). \tag{3.49}$$

At $n = 32$ for example, as in our simulations, the conventional receiver metric is around $39\times$ faster in terms of detection than the JED ($M = 4$) metric over pure LOS condition ($\alpha = 0$) in log domain versus $25\times$ against its max-log domain version. This makes the JED metric in max-log domain about $1.5\times$ faster than the log-domain version. It should also be emphasized that by also increasing M (e.g., $M = 8$), we'll expect the complexity to escalate, making the block-wise JED approach difficult to handle.

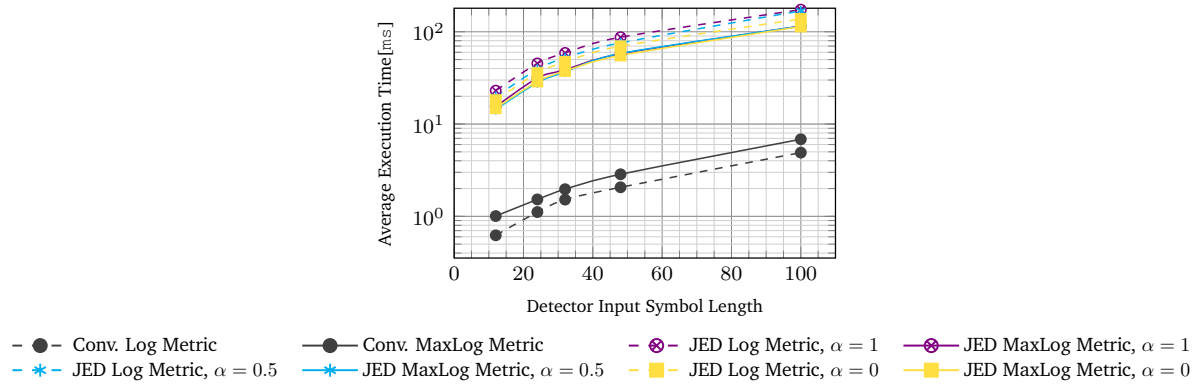


Figure 3.7: Average Time Complexity of the proposed SIMO JED metrics vs Conventional metrics.

4.3 Current Limitations and Future Work

This proposed joint detection approach has inherent limitations. It becomes computationally prohibitive for higher-order modulation schemes like 16-QAM (where $|\boldsymbol{\chi}| = 2^4$). With a joint detection block size of $M = 4$, this would require processing $|\boldsymbol{\chi}|^M = 65{,}536$ possible symbol vector combinations, which is computationally intractable. In such scenarios, reducing the detection block size (e.g., to $M = 2$) becomes necessary to maintain feasible complexity. Furthermore, the proposed JED-based receivers can be extended to incorporate iterative processing for improved LLR reliability and enhanced transmission quality. However, it should be noted that this would introduce additional receiver latency due to the extra processing layer. Hence, the iterative processes usage must therefore be motivated by specific system design requirements, complexity, available resources, transmission channel conditions, and performance goals. 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. A minimal number of iterations must be used for basic adjustments, while more complex scenarios may implement multiple iterative processes for progressive optimization. Considering recent advances in this field, one can couple this proposed metrics with the iterative estimation detection approach proposed by Jiao *et al.* [121].

5 Summary

This chapter presented enhanced BICM receiver metrics for joint estimation–detection in short blocklength transmissions, specifically targeting scenarios with unknown channel state information and limited or sparse training resources. We started with the prior art in BICM receivers, stating the general framework of the polar coded/LDPC transmission procedure and the BICM receiver design that covers non-coherent fading channels in a SIMO transmission with the aim of efficiently addressing the challenges posed by LOS and NLOS channels. We then presented the numerical results to demonstrate the merits of the propositions outlined in the chapter.

In this respect, through block-wise joint estimation–detection, we demonstrated significant improvements in performance and sensitivity compared to conventional receivers. Our analysis, conducted using full 5G transmitter and receiver chains with both Polar and LDPC coded transmissions under QPSK modulation, shows that even when reference signals are interleaved with coded data over a small number of OFDM symbols, precluding near-perfect channel estimation. Unlike conventional symbol-by-symbol detection in BICM systems, where the observation for

a given coded bit is confined to the symbol in which it is conveyed,the proposed method performs block-wise joint detection over a sliding window of adjacent symbols to fundamentally leverages their statistical dependencies; that is, the LLR for each coded bit incorporates information from multiple symbols rather than being confined to its host symbol. Performance evaluation confirms the effectiveness of this strategy in realistic base station receiver scenarios. Notably, detection windows spanning approximately four modulated symbols allow the proposed receivers to substantially outperform conventional ones, achieving detection performance approaching that of coherent receivers with perfect CSI. Overall, the presented joint estimation–detection strategy built upon an enhanced likelihood metrics offers a significant advancement in receiver design for B5G/6G, especially in highly dynamic or resource-constrained environments.

Chapter 4

Receiver Design (Part II)

Contents

1 BICM Receiver for MIMO systems

1.1 MIMO Rayleigh Block Fading Channel

Consider a SU-MIMO transmission model with N_T transmit and N_R receive antennas operating over a quasi-static, flat-fading channel. 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 $h_{t,r}$ represent the complex-valued path gain, serving as the fading coefficient from transmit antenna t to receive antenna r. At any given time instance, the complex-valued signal $\mathbf{x} = \{x_1, x_2, \ldots, x_{N_T} \mid \mathbf{x} \in \mathbb{C}^{1\times N_T}\}$ are transmitted through the N_T respective antennas. The received signal $\mathbf{y}$ belongs to the complex vector space $\mathbb{C}^{1\times N_R}$. $\mathbf{z} \sim \mathbb{C}^{1\times N_R}$ is the additive white Gaussian noise with independent real and imaginary components, each having a variance of σ^2 in every dimension, and the channel matrix is denoted by $\mathbf{H} \in \mathbb{C}^{N_T\times N_R}$. The MIMO channel model is succinctly expressed as

$$\mathbf{y} = \mathbf{x}\mathbf{H} + \mathbf{z}. \tag{4.1}$$

However, the system model given in (4.1), which describes transmission during a single symbol interval, can be extended to multiple consecutive transmissions. Consider N consecutive transmitted vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, where N denotes the total number of symbol intervals.
For notational convenience, we organize these vectors into matrices as follows: $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^\top \in \mathbb{C}^{N\times N_R}$, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]^\top \in \mathbb{C}^{N\times N_R}$, $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N]^\top \in \mathbb{C}^{N\times N_R}$. The associated block transmission model can then be compactly expressed as

$$\begin{bmatrix} y_{1,1} & \cdots & y_{1,N_R} \\ y_{2,1} & \cdots & y_{2,N_R} \\ \vdots & \ddots & \vdots \\ y_{N,1} & \cdots & y_{N,N_R} \end{bmatrix} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,N_T} \\ x_{2,1} & \cdots & x_{2,N_T} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,N_T} \end{bmatrix} \begin{bmatrix} h_{1,1} & \cdots & h_{1,N_R} \\ h_{2,1} & \cdots & h_{2,N_R} \\ \vdots & \ddots & \vdots \\ h_{N_T,1} & \cdots & h_{N_T,N_R} \end{bmatrix} + \begin{bmatrix} z_{1,1} & \cdots & z_{1,N_R} \\ z_{2,1} & \cdots & z_{2,N_R} \\ \vdots & \ddots & \vdots \\ z_{N,1} & \cdots & z_{N,N_R} \end{bmatrix}, \tag{4.2}$$

or equivalently,

$$\mathbf{Y} = \mathbf{XH} + \mathbf{Z}. \tag{4.3}$$

where $\mathbf{H}$ is assumed to be constant over the N symbol intervals or the N-symbol block and changes independently across blocks. We adopt a wide assumption regarding $\mathbf{H} \sim \mathbb{CN}(0, \mathbf{I})$, which is that its entries, $\mathrm{h}_{t,r}$, are statistically independent for the sake of simplicity. In practice, the complex path gains $|\mathrm{h}_{t,r}|$ exhibit correlations influenced by factors such as the propagation environment, antenna element polarization, and their spacing. Accordingly, the complex-valued fading coefficients $\mathrm{h}_{t,r}$ can be treated as independent zero-mean complex Gaussian random variables with unit variance. Therefore, the MIMO channel model can be referred as the *identically and independently distributed Rayleigh fading channel*, or more precisely, the Rayleigh block-fading.

Definition 4.1. (Block fading channel)
Block fading is where the fading process is approximately constant for the number of symbol interval. Notice that a channel can be double block fading when it is block fading in both time and frequency domain. Many wireless communication channels are dynamic by nature, and they are commonly modelled as block fading.

As stated earlier, 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.
Figure 4.1 depicts the transmit-end procedure for uplink channels over a $(\mathsf{N_T} \times \mathsf{N_R})$ *point-to-point* (P2P) MIMO System.

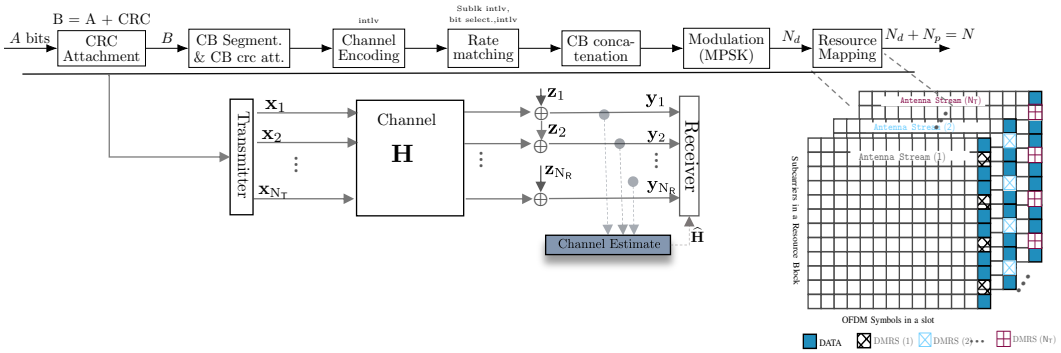


Figure 4.1: Bit-Interleaved Polar/LDPC coded Modulation (BIPCM/BILCM), a $(\mathsf{N_T} \times \mathsf{N_R})$ SU-MIMO System.

1.1.1 Signal Orthogonality in Pilot Allocation

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 exploring 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[142]. For the primary case of interest here, signal orthogonality is adopted for the MIMO configuration. Indeed, in a MIMO transmission system with $(\mathsf{N_T} \geq 1)$, these antennas often share the same time-frequency resources for DMRS. To prevent inter-antenna interference, Zadoff-Chu sequences with the same root are orthogonalized across antennas using cyclic shifts (i.e., linear phase rotations). Thus we define the DMRS sequence $\mathrm{x}_t^{(\mathsf{p})}[n]$ transmitted via antenna port $t = 1, 2, \ldots, \mathsf{N_T}$ by

$$\mathrm{x}_t^{(\mathsf{p})}[n] = \mathrm{x}^{(\mathsf{p})}[n] \cdot \exp\left(j \cdot \frac{2\pi(t-1)n}{\mathsf{N_T}}\right), \ \forall n \in \mathcal{N}_p. \tag{4.4}$$

where $\mathrm{x}^{(\mathsf{p})}[n]$ is given in (3.2) in the previous chapter. The second term $e^{i \cdot \phi_t(n)}$, where $\phi_t(n) = \frac{2\pi(t-1)n}{\mathsf{N_T}}$, applies a linear cyclic phase shift to orthogonalize the DMRS sequences across different transmit antennas. To ensure this orthogonality, as required by the 5G standard, an additional condition must be satisfied. Indeed, the number of DMRS

positions N_p must be at least equal to the number of antenna ports $\mathsf{N_T}$, which means: $\mathsf{N_T} \leq N_p$ [141].

1.1.2 DMRS Density per antenna Stream

For the MIMO case, unlike the SIMO transmission type presented earlier, we need to increase the DMRS density per PRB or per antenna stream. Simply having orthogonal sequences is not enough to achieve good channel-estimation quality. For example, in a 4×4 MIMO configuration, using only one DMRS per antenna stream is not sufficient to properly estimate all four channels—it can only identify each antenna's channel, but not estimate them reliably. Therefore, densifying the DMRS sequences (e.g., using four DMRS symbols per antenna stream) is required to ensure stable and accurate channel estimation. Since a single orthogonal DMRS sequence can uniquely identify each antenna's channel, it does not provide reliable estimation, as it remains sensitive to channel variations, noise, and interferences.

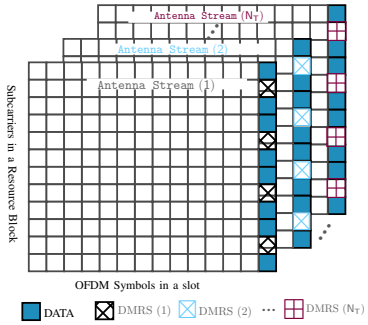


Figure 4.2: Conceptual illustration of DMRS density per stream in MIMO Configuration

1.1.3 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:

$$\mathrm{q}\left(\mathbf{X}, \{\mathbf{Y}, \mathbf{H}\}\right) \propto \exp\left(-\frac{1}{\mathrm{N_0}} \operatorname{tr}\left\{\left(\mathbf{Y} - \mathbf{XH}\right)\left(\mathbf{Y} - \mathbf{XH}\right)^{\dagger}\right\}\right). \tag{4.5}$$

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

$$\mathrm{q}\left(\mathbf{X}, \{\mathbf{Y}, \mathbf{H}\}\right) \propto \exp\left(\frac{2}{\mathrm{N_0}} \mathrm{Re}\left(\operatorname{tr}\left\{\mathbf{YH}^{\dagger}\mathbf{X}^{\dagger}\right\}\right) - \frac{1}{\mathrm{N_0}} \|\mathbf{XH}\|_{\mathsf{F}}^2\right). \tag{4.6}$$

The likelihood of the encoded bit e_j s.t. $b \in \{0, 1\}$ is defined by

$$\mathrm{q}\left(e_j(\mathbf{X}) = b, \{\mathbf{Y}, \mathbf{H}\}\right) = \sum_{\mathbf{x} \in \boldsymbol{\chi}_b^j} \mathrm{q}\left(\mathbf{X}, \{\mathbf{Y}, \mathbf{H}\}\right). \tag{4.7}$$

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

$$\Lambda^j\left(\mathbf{Y}\right) = \log \frac{\mathrm{q}\left(e_j(\mathbf{X}) = 0, \{\mathbf{Y}, \mathbf{H}\}\right)}{\mathrm{q}\left(e_j(\mathbf{X}) = 1, \{\mathbf{Y}, \mathbf{H}\}\right)}. \tag{4.8}$$

and the corresponding max-log approximation of (4.8) is shown to be

$$\Lambda^j\left(\mathbf{Y}\right) = \sum_{b \in \{0,1\}} (-1)^b \left[\max_{\mathbf{x} \in \boldsymbol{\chi}_b^j} \frac{1}{\mathrm{N_0}} \left(2\,\mathrm{Re}\left(\operatorname{tr}\left\{\mathbf{YH}^{\dagger}\mathbf{X}^{\dagger}\right\}\right) - \|\mathbf{XH}\|_{\mathsf{F}}^2\right)\right]. \tag{4.9}$$

Remark 4.1. This metric is typically used in MIMO receivers based on Perfect CSI as well as in conventional quasi-coherent MIMO receivers, and both apply so-called "symbol-by-symbol" detection, meaning that each symbol is detected independently of the other symbols.

In other respects, the conventional receiver is obtained by replacing $\mathbf{H}$ with the estimated channel matrix $\hat{\mathbf{H}}$ via the separate least squares method. The separate channel estimation by the least squares method aims to independently estimate the channel characteristics for each antenna, ignoring spatial correlation. More precisely, for the antenna

pair (t, r), the channel matrix element is computed as follows: $\hat{\mathbf{h}}_{t,r} = \frac{\mathbf{x}_t^\dagger \mathbf{y}_r}{\mathbf{x}_t^\dagger \mathbf{x}_t}$. This method is often used in scenarios where the antennas of the MIMO system are sufficiently isolated from each other and where there is no correlation between the different antennas.
We consider these receivers as a benchmark for comparison with respect to the subsequent proposed advanced receivers

1.1.4 No Channel State Information

As stated earlier, under the block-fading assumption, each of the complex-valued fading coefficient $\mathrm{h}_{t,r}$ is assumed to remain constant over the N symbol periods and to be independent across blocks. Therefore, their probability density function can be expressed as follows $\mathrm{p}\left(\mathrm{h}_{t,r}\right) = \frac{1}{\pi} \exp\left(-|\mathrm{h}_{t,r}|^2\right)$. 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 4.2. (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:

$$f_{\boldsymbol{x}}(\boldsymbol{v}) \triangleq \frac{1}{\pi \det\left(\mathbf{\Phi}_{\boldsymbol{x}}\right)} \exp\left[-\left(\boldsymbol{v} - \boldsymbol{\mu}_{\boldsymbol{x}}\right)^\dagger \mathbf{\Phi}_{\boldsymbol{x}}^{-1} \left(\boldsymbol{v} - \boldsymbol{\mu}_{\boldsymbol{x}}\right)\right] . \tag{4.10}$$

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

$$\mathrm{p}\left(\mathbf{Y}|\mathbf{X}\right) = \frac{\exp\left(-\operatorname{tr}\left\{\mathbf{Y}\mathbf{\Phi}_{\mathsf{Y}}^{-1}\mathbf{Y}^\dagger\right\}\right)}{\pi^{N \times \mathsf{N_R}} \det^{\mathsf{N_R}}\left(\mathbf{\Phi}_{\mathsf{Y}}\right)} . \tag{4.11}$$

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

Theorem 4.1. [136, 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:

$$\begin{aligned}
\mathbf{\Phi}_{\mathsf{Y}} &\triangleq \mathbb{E}\{\mathbf{YY}^\dagger\}, \\
&= \mathbb{E}\{\left(\mathbf{XH} + \mathbf{Z}\right)\left(\mathbf{XH} + \mathbf{Z}\right)^\dagger\}, \\
&= \mathbb{E}\{\mathbf{XHH}^\dagger\mathbf{X}^\dagger\} + \mathbb{E}\{\mathbf{ZZ}^\dagger\}, \\
&= \mathbf{X}\,\mathbb{E}\{\mathbf{HH}^\dagger\}\,\mathbf{X}^\dagger + \mathbb{E}\{\mathbf{ZZ}^\dagger\}, \\
&= \mathbf{X}\,\mathbf{I}\,\mathbf{X}^\dagger + \mathbb{E}\{\mathbf{ZZ}^\dagger\}, \\
&= \mathbf{XX}^\dagger + 2\sigma^2\mathbf{I} = \mathbf{XX}^\dagger + \mathrm{N}_0\mathbf{I} .
\end{aligned} \tag{4.12}$$

This expression for the covariance matrix is commonly encountered in the literature[90, 91].
Let's now introduce some relevant mathematical properties, which we shall consider in the metric derivation steps.

Property 4.1. (Some useful formulas)

1. $\det\left(\mathbf{I} + \mathbf{AB}\right) = \det\left(\mathbf{I} + \mathbf{BA}\right)$ [137],
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:

$$\det \mathbf{\Phi}_{\mathsf{Y}} = \det\left(\mathrm{N}_0\mathbf{I} + \mathbf{X}\mathbf{X}^\dagger\right) \tag{4.13}$$

$$= (\lambda_1 + \mathrm{N}_0)(\lambda_2 + \mathrm{N}_0)\cdots(\lambda_N + \mathrm{N}_0), \quad \text{where} \quad \mathrm{Eigenvalue}(\mathbf{X}\mathbf{X}^\dagger) = \lambda_n, n = 1, 2, \ldots N. \tag{4.14}$$

$$= \prod_{n=1}^{N} (\lambda_n + \mathrm{N}_0)\,. \tag{4.15}$$

Proof. (Determinant of $\mathbf{\Phi}_{\mathsf{Y}}$.)
To compute the determinant of $\mathbf{\Phi}_{\mathsf{Y}}$, consider the eigenvalue decomposition of the Hermitian matrix: $\mathbf{X}\mathbf{X}^\dagger = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$, where $\mathbf{U}$ is a unitary matrix, and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$ is a diagonal matrix of eigenvalues λ_n ($n = 1, \ldots, N$). Using this decomposition, we can write:

$$\mathrm{N}_0\mathbf{I} + \mathbf{X}\mathbf{X}^\dagger = \mathbf{U}\left(\mathrm{N}_0\mathbf{I} + \mathbf{\Lambda}\right)\mathbf{U}^\dagger. \tag{4.16}$$

Since the determinant of a unitary matrix is 1, the determinant simplifies to:

$$\det\left(\mathrm{N}_0\mathbf{I} + \mathbf{X}\mathbf{X}^\dagger\right) = \det\left(\mathrm{N}_0\mathbf{I} + \mathbf{\Lambda}\right), \tag{4.17}$$

and for the diagonal matrix $\mathrm{N}_0\mathbf{I} + \mathbf{\Lambda}$, the determinant is the product of its diagonal elements.

$$\det\left(\mathrm{N}_0\mathbf{I} + \mathbf{X}\mathbf{X}^\dagger\right) = \prod_{n=1}^{N}\left(\mathrm{N}_0 + \lambda_n\right), \tag{4.18}$$

where λ_n are the eigenvalues of $\mathbf{X}\mathbf{X}^\dagger$. If $\mathbf{X}$ is rank-deficient ($\mathrm{rank}(\mathbf{X}) = r < N$), only the first r eigenvalues are non-zero. ■

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 4.1. [138, The Woodbury Matrix identity].

$$(\mathbf{A} + \mathbf{U}\mathbf{C}\mathbf{V})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{C}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}\mathbf{A}^{-1}, \tag{4.19}$$

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: $\mathbf{\Phi}_{\mathsf{Y}}^{-1} = \left(\mathrm{N}_0\mathbf{I} + \mathbf{X}\mathbf{X}^\dagger\right)^{-1}$.

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,

$$\begin{aligned}\mathbf{\Phi}^{-1} &= (\mathbf{A} + \mathbf{U}\mathbf{C}\mathbf{V})^{-1}\,,\\ &= \mathrm{N}_0^{-1}\mathbf{I} - \mathrm{N}_0^{-1}\mathbf{X}\left[\mathrm{N}_0\mathbf{I} + \mathbf{X}^\dagger\mathbf{X}\right]^{-1}\mathbf{X}^\dagger,\\ &= \mathrm{N}_0^{-1}\mathbf{I} - \mathrm{N}_0^{-1}\mathbf{X}\mathbf{D}\mathbf{X}^\dagger,\end{aligned} \tag{4.20}$$

where

$$\mathbf{D} = \left[\mathrm{N}_0\mathbf{I} + \mathbf{X}^\dagger\mathbf{X}\right]^{-1}. \tag{4.21}$$

Property 4.2. (Matrix $\mathbf{D}$ decomposition)
Assuming that $\mathbf{X}$ is a matrix of size $N \times \mathsf{N_T}$, and thus $\mathbf{X}^\dagger\mathbf{X}$ is a square matrix of size $\mathsf{N_T} \times \mathsf{N_T}$.
We can then decompose the expression of $\mathbf{D}$ using eigenvalue decomposition (EVD) or, more generally, singular value decomposition (SVD).

1. **Eigenvalue decomposition:**
   Let the eigenvalue decomposition of the Hermitian matrix be expressed as:

$$\mathbf{X}^\dagger\mathbf{X} = \{\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger \quad | \quad \mathbf{U}^\dagger\mathbf{U} = \mathbf{I}, \mathbf{U}^\dagger = \mathbf{U}^{-1}, |\det(\mathbf{U})| = 1\} \tag{4.22}$$

where $\mathbf{U}$ is a unitary matrix of size $\mathsf{N_T} \times \mathsf{N_T}$ and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{\mathsf{N_T}})$ is a diagonal matrix containing the eigenvalues λ_i of Hermitian matrix.

Substituting this decomposition into the expression for $\mathbf{D}$, we get: $\mathbf{D} = \left[\mathrm{N}_0\mathbf{I} + \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger\right]^{-1}$.

Since $\mathbf{U}$ is a unitary matrix, we can simplify the inverse as:$\mathbf{D} = \mathbf{U}\left[\mathrm{N}_0\mathbf{I} + \mathbf{\Lambda}\right]^{-1}\mathbf{U}^\dagger$. Finally, the inverse of the matrix $\mathrm{N}_0\mathbf{I} + \mathbf{\Lambda}$, which is diagonal, is computed element-wise:

$$\left[\mathrm{N}_0\mathbf{I} + \mathbf{\Lambda}\right]^{-1} = \text{diag}\left(\frac{1}{\mathrm{N}_0 + \lambda_1}, \frac{1}{\mathrm{N}_0 + \lambda_2}, \ldots, \frac{1}{\mathrm{N}_0 + \lambda_{\mathsf{N_T}}}\right). \tag{4.23}$$

Thus, the simplified form of $\mathbf{D}$ is given by:

$$\mathbf{D} = \mathbf{U}\,\text{diag}\left(\frac{1}{\mathrm{N}_0 + \lambda_1}, \frac{1}{\mathrm{N}_0 + \lambda_2}, \ldots, \frac{1}{\mathrm{N}_0 + \lambda_{\mathsf{N_T}}}\right)\mathbf{U}^\dagger. \tag{4.24}$$

However, the use of eigenvalue decomposition is not always possible or appropriate in certain cases. Indeed, eigenvalue decomposition is typically employed for *Hermitian matrices* (i.e., $[\mathbf{X}^\dagger\mathbf{X}] = [\mathbf{X}^\dagger\mathbf{X}]^\dagger$) or *symmetric matrices*. Conversely, if the matrix $\mathbf{X}^\dagger\mathbf{X}$ is not Hermitian (which may occur if $\mathbf{X}$ is not a real-valued matrix or lacks a particular structure), eigenvalue decomposition may not be directly applicable in a standard manner. In such cases, the use of *singular value decomposition* (SVD), a more general approach that works for all matrices, should be considered.

2. **Singular Value Decomposition (SVD) for the Matrix $\mathbf{D}$:**

In this case, we assume that the matrix $\mathbf{X}$ has complex coefficients, but it is not necessarily Hermitian. Therefore, we need to use the SVD to handle the matrix properly, even when it is not square or well-conditioned. Here, in particular, $\mathbf{X}^\dagger \in \mathbb{C}^{N \times \mathsf{N_T}}$ matrix, which may not be well-conditioned, especially if $\mathbf{X}$ has a low rank.

Since the SVD is applicable to any complex matrix, even if it is not square or Hermitian, then SVD of $\mathbf{X}$ is of the form:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger, \tag{4.25}$$

where: $\mathbf{U} \in \mathbb{C}^{N \times N}$ is a unitary matrix, $\mathbf{V} \in \mathbb{C}^{\mathsf{N_T} \times \mathsf{N_T}}$ is a unitary matrix, and $\mathbf{\Sigma} \in \mathbb{C}^{N \times \mathsf{N_T}}$ is a diagonal (rectangular) matrix whose nonzero diagonal entries representing the singular values $\left(\sigma_i\right)_{i=0}^{r}$ of $\mathbf{X}$. Hence we can express $\mathbf{X}^\dagger\mathbf{X}$ as:

$$\mathbf{X}^\dagger\mathbf{X} = \mathbf{V}\mathbf{\Sigma}^\dagger\underbrace{\mathbf{U}^\dagger\mathbf{U}}_{=\mathbf{I}}\mathbf{\Sigma}\mathbf{V}^\dagger = \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{\Sigma}\mathbf{V}^\dagger. \tag{4.26}$$

This decomposition is useful for studying the matrix $\mathbf{X}^\dagger\mathbf{X}$, even when $\mathbf{X}$ is neither square nor Hermitian.

Using the SVD, the matrix $\mathbf{D}$ becomes: $\mathbf{D} = \left[\mathrm{N}_0\mathbf{I} + \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{\Sigma}\mathbf{V}^\dagger\right]^{-1}$, and since $\mathbf{V}$ is a unitary matrix, we can simplify this expression as: $\mathbf{D} = \mathbf{V}\left[\mathrm{N}_0\mathbf{I} + \mathbf{\Sigma}^\dagger\mathbf{\Sigma}\right]^{-1}\mathbf{V}^\dagger$.

Since $\mathbf{\Sigma}^\dagger\mathbf{\Sigma}$ is a diagonal matrix containing the squares of the singular values of $\mathbf{X}$, its inverse is also a diagonal matrix:

$$\left[\mathrm{N}_0\mathbf{I} + \mathbf{\Sigma}^\dagger\mathbf{\Sigma}\right]^{-1} = \text{diag}\left(\frac{1}{\mathrm{N}_0 + \sigma_1^2}, \frac{1}{\mathrm{N}_0 + \sigma_2^2}, \ldots, \frac{1}{\mathrm{N}_0 + \sigma_r^2}\right). \tag{4.27}$$

Substituting the inverse, the final form of $\mathbf{D}$ is:

$$\mathbf{D} = \mathbf{V}\,\text{diag}\left(\frac{1}{\mathrm{N}_0 + \sigma_1^2}, \frac{1}{\mathrm{N}_0 + \sigma_2^2}, \ldots, \frac{1}{\mathrm{N}_0 + \sigma_r^2}\right)\mathbf{V}^\dagger, \text{where } r = \text{rank}(\mathbf{X}) \leq \min\left(N, \mathsf{N_T}\right). \tag{4.28}$$

Note that the singular values $\sigma_i, \forall i = 1, 2, \ldots, r$ are the square roots of the eigenvalues of $\mathbf{X}^\dagger\mathbf{X}$ (i.e., $\sigma_i = \sqrt{\lambda_i(\mathbf{X}^\dagger\mathbf{X})}$. The singular values are always real and non-negative(i.e., $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$) and they somewhat play a key role in numerical stability.

Thus, the likelihood function can be straightforwardly formulated as follows:

$$\begin{aligned}
\mathrm{p}\left(\mathbf{Y}|\mathbf{X}\right) &= \frac{1}{\mathbf{L}_{\mathsf{X}}} \exp\left(-\operatorname{tr}\left\{\mathbf{Y}^{\dagger}\boldsymbol{\Phi}_{\mathsf{Y}}^{-1}\mathbf{Y}\right\}\right) \\
&= \frac{1}{\mathbf{L}_{\mathsf{X}}} \exp\left(-\operatorname{tr}\left\{\mathbf{Y}^{\dagger}\left(\frac{1}{\mathrm{N}_0}\mathbf{I} - \frac{1}{\mathrm{N}_0}\mathbf{X}\mathbf{D}\mathbf{X}^{\dagger}\right)\mathbf{Y}\right\}\right), \\
&= \frac{1}{\mathbf{L}_{\mathsf{X}}} \exp\left(-\frac{1}{\mathrm{N}_0}\operatorname{tr}\left\{\mathbf{Y}^{\dagger}\mathbf{Y}\right\} + \frac{1}{\mathrm{N}_0}\operatorname{tr}\left\{\mathbf{Y}^{\dagger}\mathbf{X}\mathbf{D}\mathbf{X}^{\dagger}\mathbf{Y}\right\}\right), \\
&= \frac{1}{\mathbf{L}_{\mathsf{X}}} \exp\left(-\frac{1}{\mathrm{N}_0}\operatorname{tr}\left\{\mathbf{Y}^{\dagger}\mathbf{Y}\right\}\right) \exp\left(\frac{1}{\mathrm{N}_0}\operatorname{tr}\left\{\left(\mathbf{X}^{\dagger}\mathbf{Y}\right)^{\dagger}\mathbf{D}\mathbf{X}^{\dagger}\mathbf{Y}\right\}\right),
\end{aligned} \tag{4.29}$$

where $\mathbf{L}_{\mathsf{X}} = \pi^{N\times \mathsf{N}_{\mathsf{R}}} \det^{\mathsf{N}_{\mathsf{R}}}\left(\mathrm{N}_0\mathbf{I} + \mathbf{X}\mathbf{X}^{\dagger}\right)$.
We ignore the multiplicative terms independent of $\mathbf{X}$ in (4.29).

Proposition 4.1. (A novel soft-likelihood metric for MIMO.)
The proposed likelihood function $\mathrm{q}\left(\mathbf{X},\mathbf{Y}\right) = p\left(\mathbf{Y}|\mathbf{X}\right)$ can be stated as follows:

$$\mathrm{q}\left(\mathbf{X},\mathbf{Y}\right) \propto \frac{1}{\mathbf{L}_{\mathsf{X}}} \exp\left(\frac{1}{\mathrm{N}_0}\operatorname{tr}\left\{\left(\mathbf{X}^{\dagger}\mathbf{Y}\right)^{\dagger}\mathbf{D}\left(\mathbf{X}^{\dagger}\mathbf{Y}\right)\right\}\right). \tag{4.30}$$

1.1.5 Joint Estimation and Detection for MIMO

As described in the proposition 3.2, 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:

$$\begin{aligned}
\mathbf{X}^{\dagger}\mathbf{Y} &= \underbrace{\mathbf{X}^{(\mathsf{p})\dagger}\mathbf{Y}^{(\mathsf{p})}}_{\text{channel estimate}} + \mathbf{X}^{(\mathsf{d})\dagger}\mathbf{Y}^{(\mathsf{d})} \\
&= \mathbf{C}_{\mathsf{p}}\widehat{\mathbf{H}}_{\mathrm{LS}} + \mathbf{X}^{(\mathsf{d})\dagger}\mathbf{Y}^{(\mathsf{d})},
\end{aligned} \tag{4.31}$$

where $\mathbf{C}_{\mathsf{p}} = \mathbf{X}^{(\mathsf{p})\dagger}\mathbf{X}^{(\mathsf{p})}$ given that $\widehat{\mathbf{H}}_{\mathrm{LS}} = \frac{\mathbf{X}^{(\mathsf{p})\dagger}\mathbf{Y}^{(\mathsf{p})}}{\mathbf{X}^{(\mathsf{p})\dagger}\mathbf{X}^{(\mathsf{p})}}$.
The channel estimate is obtained by jointly estimating the channel using the least squares (LS) method and by averaging over the number of dimensions representing a coherent channel. The joint least squares method enables the consideration of spatial correlation, resulting in a more accurate channel estimate.

Proposition 4.2. (Joint estimation-detection MIMO metric)
Consistent with proposition 4.1, the expression of likelihood function that enables joint estimation and detection in MIMO scenario, can be formulated as follows:

$$\mathrm{q}(\mathbf{X},\mathbf{Y}) = \frac{1}{\mathbf{L}_{\mathsf{X}}} \exp\left(\frac{1}{\mathrm{N}_0}\operatorname{tr}\left\{\left(\mathbf{C}_{\mathsf{p}}\widehat{\mathbf{H}}_{\mathrm{LS}} + \mathbf{X}^{(\mathsf{d})\dagger}\mathbf{Y}^{(\mathsf{d})}\right)^{\dagger}\mathbf{D}\left(\mathbf{C}_{\mathsf{p}}\widehat{\mathbf{H}}_{\mathrm{LS}} + \mathbf{X}^{(\mathsf{d})\dagger}\mathbf{Y}^{(\mathsf{d})}\right)\right\}\right). \tag{4.32}$$

The likelihood of the encoded bit e_j s.t. $b \in \{0,1\}$ is defined by

$$\mathrm{q}\left(e_j(\mathbf{X}) = b, \mathbf{Y}\right) = \sum_{\mathbf{x}\,\in\,\boldsymbol{\chi}_b^j} \mathrm{q}\left(\mathbf{X},\mathbf{Y}\right). \tag{4.33}$$

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

$$\Lambda^j\left(\mathbf{Y}\right) = \log\frac{\mathrm{q}\left(e_j(\mathbf{X}) = 0, \mathbf{Y}\right)}{\mathrm{q}\left(e_j(\mathbf{X}) = 1, \mathbf{Y}\right)}. \tag{4.34}$$

Corollary 4.1. To ease the process of implementing such a LLR bit metric in (4.34), one may use its *max-log approximation* version given in (4.35).

$$\Lambda^j(\mathbf{Y}) \;=\; \sum_{b\in\{0,1\}} (-1)^b \left[\max_{\mathbf{X}\in\boldsymbol{\chi}_b^j} \Gamma\big(\mathbf{X},\mathbf{Y}\big) \;-\; \sum_{\mathbf{X}\in\boldsymbol{\chi}_b^j} \log \mathbf{L}_{\mathsf{X}}\right]. \tag{4.35}$$

with the score function or max-log decision metric

$$\Gamma\big(\mathbf{X},\mathbf{Y}\big) \triangleq \frac{1}{\mathrm{N}_0}\ \mathrm{tr}\Big\{ \big(\mathbf{C}_p\,\widehat{\mathbf{H}}_{\mathrm{LS}} + \mathbf{X}^{(d)\,\dagger}\,\mathbf{Y}^{(d)}\big)^{\dagger}\ \mathbf{D}\ \big(\mathbf{C}_p\,\widehat{\mathbf{H}}_{\mathrm{LS}} + \mathbf{X}^{(d)\,\dagger}\,\mathbf{Y}^{(d)}\big)\Big\}. \tag{4.36}$$

Remark 4.2. 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 (3.12) 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.

1.1.6 Block-wise Joint Estimation and Detection

For a MIMO system with $\mathsf{N_T}$ transmit antennas employing block-wise detection (block size M) and QPSK modulation ($|\boldsymbol{\chi}| = 4$), each antenna's symbol vector is selected from $|\boldsymbol{\chi}|^M$ possible candidates. The joint search space across all $\mathsf{N_T}$ antennas then expands to $\left(|\boldsymbol{\chi}|^M\right)^{\mathsf{N_T}}$ possible combinations - computationally expensive but optimal (ML joint detection). However, under the assumption of no spatial correlation, the search space can be processed independently for each of the $\mathsf{N_T}$ streams. In this case, each transmit stream is detected independently without inter-antenna joint processing. This stream-wise implementation reduces the computational complexity from exponential to linear scaling with respectt to $\mathsf{N_T}$, requiring only $\mathsf{N_T} \times |\boldsymbol{\chi}|^M$ total candidates. This per-stream block-wise joint detection approach demonstrates somewhat significantly reduced complexity compared to full joint detection (i.e., joint-stream block-wise detection).
Algorithm 2 outlines this stream-wise implementation of block-wise JED for spatially multiplexed MIMO systems.

Algorithm 3: Stream-wise Block-wise Joint Detection (MIMO).

Input: Received frame: $\mathbf{Y} \in \mathbb{C}^{N \times \mathsf{N_R}}$, Data and DMRS indices: (d), (p), Candidate list size per stream: $L = |\boldsymbol{\chi}|^M$, Candidate symbol tensor: $\mathbf{X}^{(\mathsf{d})} \in \boldsymbol{\chi}^{L \times M \times \mathsf{N_T}}$, where each candidate block $\mathbf{X}^{(\mathsf{d})}(c,:,t) \in \boldsymbol{\chi}^M$, Noise power: N_0

Output: LLR vector for all coded bits: $\boldsymbol{\Lambda}(\mathbf{Y}^{(\mathsf{d})})$

begin

Step 1: Joint Channel Estimation (LS)

Extract DMRS-based LS estimate for each transmit antenna $t = 1, \ldots, \mathsf{N_T}$:

$$\widehat{\mathbf{H}}_{\mathrm{LS}}(t,:) = \frac{\left[\mathbf{X}^{(\mathsf{p})}(:,t)\right]^\dagger \mathbf{Y}^{(\mathsf{p})}}{\left[\mathbf{X}^{(\mathsf{p})}(:,t)\right]^\dagger \left[\mathbf{X}^{(\mathsf{p})}(:,t)\right]}, \quad \widehat{\mathbf{H}}_{\mathrm{CP}}(t,:) = \left[\mathbf{X}^{(\mathsf{p})}(:,t)\right]^\dagger \mathbf{Y}^{(\mathsf{p})}.$$

Compute MMSE equalizer once:

$$\mathbf{W} = \left(\widehat{\mathbf{H}}_{\mathrm{LS}}^\dagger \widehat{\mathbf{H}}_{\mathrm{LS}} + \mathrm{N}_0\, \mathbf{I}_{\mathsf{N_T}}\right)^{-1} \widehat{\mathbf{H}}_{\mathrm{LS}}^\dagger.$$

Step 2: Stream-wise Block Processing

for *each transmit stream* $t = 1, \ldots, \mathsf{N_T}$ **do**

- Extract equalizer vector $\mathbf{w}_t = [\mathbf{W}(t,:)]^\dagger$,
- and compute effective noise: $\mathrm{N}_{0,\mathrm{eff}}^{(t)} = \mathrm{N}_0 \left\|\mathbf{w}_t\right\|^2$.
- *For each block* $b = 1, \ldots, B$:
- Compute block-index range and extract data block:
- Set $\mathsf{idx}_b = [(b-1)\,M + 1 : b\,M]$, then $\mathbf{Y}_b^{(\mathsf{d})} = \mathbf{Y}^{(\mathsf{d})}(\mathsf{idx}_b, :)$.
- Equalize data block : $\widetilde{\mathbf{X}}_b^{(\mathsf{d})} = \mathbf{Y}_b^{(\mathsf{d})}\, \mathbf{W}$ (size: $M \times \mathsf{N_T}$).
- Extract stream-t component and form intermediate observation:
- $\mathbf{Y}_{b,t}^{(\mathsf{d})} = \widetilde{\mathbf{X}}_b^{(\mathsf{d})}(:,t)\, [\widehat{\mathbf{H}}_{\mathrm{LS}}(t,:)] \in \mathbb{C}^{M \times \mathsf{N_R}}$.
- *Step 2.1: Likelihood Function Evaluation*
- **for** *each candidate index* $c = 1, \ldots, L$ **do**
  - Let $\mathbf{X}_c^{(\mathsf{d})}(:,t) \triangleq \mathbf{X}^{(\mathsf{d})}(c,:,t) \in \boldsymbol{\chi}^{M \times 1}$.
  - Compute: $\mathbf{A}(t,:) = \widehat{\mathbf{H}}_{\mathrm{CP}}(t,:) + \left[\mathbf{X}_c^{(\mathsf{d})}(:,t)\right]^\dagger \mathbf{Y}_{b,t}^{(\mathsf{d})}$.
  - Compute $\mathbf{D}(t,t) = \left(\mathrm{N}_{0,\mathrm{eff}}^{(t)} + \left[\mathbf{X}_c^{(\mathsf{d})}(:,t)\right]^\dagger \left[\mathbf{X}_c^{(\mathsf{d})}(:,t)\right]\right)^{-1}$.
  - Compute the metric core function and Stabilize:
  - $\Gamma(\mathbf{X}_c^{(\mathsf{d})}(:,t),\, \mathbf{Y}_{b,t}^{(\mathsf{d})}) = \frac{1}{\mathrm{N}_0} \sum_{r=1}^{\mathsf{N_R}} \mathbf{A}^\dagger(t,r)\, \mathbf{D}(t,t)\, \mathbf{A}(t,r)$.
  - Evaluate likelihood metric:
  - $\mathrm{q}(\mathbf{X}_c^{(\mathsf{d})}(:,t), \mathbf{Y}_{b,t}^{(\mathsf{d})}) = \frac{\exp\left(\Gamma(\mathbf{X}_c^{(\mathsf{d})}(:,t), \mathbf{Y}_{b,t}^{(\mathsf{d})})\right)}{\pi^{M\mathsf{N_R}}\, \det\left(\mathrm{N}_{0,\mathrm{eff}}^{(t)}\, \mathbf{I}_M + \left[\mathbf{X}_c^{(\mathsf{d})}(:,t)\right]\, \left[\mathbf{X}_c^{(\mathsf{d})}(:,t)\right]^\dagger\right)^{\mathsf{N_R}}}$.
- *Step 2.2: LLR Computation*
- **for** *each bit index* $j = 1, \ldots, M\, \log_2 |\boldsymbol{\chi}|$ *in block* b **do**
  - Compute the LLR of coded bit $e_j \in \{0,1\}$:
  - $\Lambda_j^{(t,b)} = \log \frac{\sum_{\mathbf{X}_c^{(\mathsf{d})}(:,t) \in \boldsymbol{\chi}_0^j} \mathrm{q}(\mathbf{X}_c^{(\mathsf{d})}(:,t), \mathbf{Y}_{b,t}^{(\mathsf{d})})}{\sum_{\mathbf{X}_c^{(\mathsf{d})}(:,t) \in \boldsymbol{\chi}_1^j} \mathrm{q}(\mathbf{X}_c^{(\mathsf{d})}(:,t), \mathbf{Y}_{b,t}^{(\mathsf{d})})}$.
- Store $\{\Lambda^{(t,b)}\}$ as the LLR vector for stream t, block b.

Step 3: Concatenate All LLRs

$$\boldsymbol{\Lambda}(\mathbf{Y}^{(\mathsf{d})}) = [\underbrace{\Lambda^{(1,1)}, \ldots, \Lambda^{(1,B)}}_{\boldsymbol{\Lambda}^{(1)}},\ \underbrace{\Lambda^{(2,1)}, \ldots, \Lambda^{(2,B)}}_{\boldsymbol{\Lambda}^{(2)}},\ \ldots,\ \underbrace{\Lambda^{(\mathsf{N_T},1)}, \ldots, \Lambda^{(\mathsf{N_T},B)}}_{\boldsymbol{\Lambda}^{(\mathsf{N_T})}}]$$

return final LLR vector $\boldsymbol{\Lambda}(\mathbf{Y}^{(\mathsf{d})})$

In other respects, the computational complexity of the block-wise JED approach is typically on the order $\mathcal{O}\left(\mathsf{N_R} \times \mathsf{N_T} \times B \cdot |\boldsymbol{\chi}|^M \times M \cdot \log_2 |\boldsymbol{\chi}|\right)$.

1.2 $(\mathsf{N_R} \times 2)$ MIMO Line-Of-Sight Channel

For the sake of simplicity, let's consider a $(\mathsf{N_R} \times 2)$ MIMO configuration, given the inherent complexity in deriving metrics for spatial dimensions $\mathsf{N_T} > 2$ in 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:

$$\mathbf{y}_i = \mathrm{h}_{i,1}\mathbf{x}_1 + \mathrm{h}_{i,2}\mathbf{x}_2 + \mathbf{z}_i, \quad i = 1, 2 \ldots, \mathsf{N_R}. \tag{4.37}$$

The system model can be expanded as follows:

$$\begin{cases} \mathbf{y}_1 = \mathrm{h}_{1,1}\mathbf{x}_1 + \mathrm{h}_{1,2}\mathbf{x}_2 + \mathbf{z}_1 \\ \mathbf{y}_2 = \mathrm{h}_{2,1}\mathbf{x}_2 + \mathrm{h}_{2,2}\mathbf{x}_2 + \mathbf{z}_2 \\ \quad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\ \mathbf{y}_{\mathsf{N_R}} = \mathrm{h}_{\mathsf{N_R},1} + \mathrm{h}_{\mathsf{N_R},2}\mathbf{x}_2 + \mathbf{z}_{\mathsf{N_R}}. \end{cases} \tag{4.38}$$

The LOS channel matrix is considered to be :

$$\mathbf{H} = \begin{bmatrix} e^{j\theta_{1,1}} & e^{j\theta_{1,2}} \\ e^{j\theta_{2,1}} & e^{j\theta_{2,2}} \\ \vdots & \vdots \\ e^{j\theta_{\mathsf{N_R},1}} & e^{j\theta_{\mathsf{N_R},2}} \end{bmatrix}. \tag{4.39}$$

where $\mathbf{y}_i \sim \mathbb{C}^{N\times i}$, $(\mathbf{x}_i)_{i=1}^2, \sim \mathbb{C}^{N\times 1}$ and $\mathbf{H} \sim \mathbb{C}^{N_\mathsf{R}\times 2}$.
Explicitly, the receiver signal can simply be modelled as

$$\mathbf{y}_i = e^{j\theta_{i,1}}\mathbf{x}_1 + e^{j\theta_{i,2}}\mathbf{x}_2 + \mathbf{z}_i, \quad i = 1, 2, \ldots, N_\mathsf{R}. \tag{4.40}$$

Following the same principle as in the previous section, we can establish the conditional probability density to determine the BICM metric. It is worth noting that $\theta_{i,1}$ and $\theta_{i,2}$ are unknown to the receiver and are assumed to be i.i.d. uniformly distributed random variables over $[0, 2\pi)$. If we ignore the independent multiplicative terms of the transmitted message, the likelihood function is shown to be:

$$q\left(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{y}_i\right) \propto \int_{\theta_{i,1}} \int_{\theta_{i,2}} \exp\left(-\frac{1}{\mathrm{N}_0}\left\|\mathbf{y}_i - e^{j\theta_{i,1}}\mathbf{x}_1 - e^{j\theta_{i,2}}\mathbf{x}_2\right\|^2\right) \mathrm{d}\theta_{i,2}\mathrm{d}\theta_{i,1}. \tag{4.41}$$

By expanding the ℓ^2-norms term constituting the expression of the conditional density probability,

$$\begin{aligned} \left\|\mathbf{y}_i - e^{j\theta_{i,1}}\mathbf{x}_1 - e^{j\theta_{i,2}}\mathbf{x}_2\right\|^2 &= \left\|\mathbf{y}_i - \left(e^{j\theta_{i,1}}\mathbf{x}_1 + e^{j\theta_{i,2}}\mathbf{x}_2\right)\right\|^2, \\ &= \|\mathbf{y}_i\|^2 + \|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 + 2\,\mathrm{Re}\left(e^{j(\theta_{i,1}-\theta_{i,2})}\mathbf{x}_1\mathbf{x}_2^\dagger\right) \\ &\quad - 2\,\mathrm{Re}\left(e^{-j\theta_{i,1}}\mathbf{x}_1^\dagger\mathbf{y}_i\right) - 2\,\mathrm{Re}\left(e^{-j\theta_{i,2}}\mathbf{x}_2^\dagger\mathbf{y}_i\right), \end{aligned} \tag{4.42}$$

saying

$$\mathbf{x}_1\mathbf{x}_2^\dagger = \left|\mathbf{x}_1\mathbf{x}_2^\dagger\right| e^{j\angle\mathbf{x}_1\mathbf{x}_2^\dagger}, \quad \mathbf{x}_1^\dagger\mathbf{y}_i = \left|\mathbf{x}_1^\dagger\mathbf{y}_i\right| e^{j\angle\mathbf{x}_1^\dagger\mathbf{y}_i}, \quad \mathbf{x}_2^\dagger\mathbf{y}_i = \left|\mathbf{x}_2^\dagger\mathbf{y}_i\right| e^{j\angle\mathbf{x}_2^\dagger\mathbf{y}_i} \tag{4.43}$$

and subsequently ignoring the independent terms of $\mathbf{x}_1$ and $\mathbf{x}_2$, this lends to :

$$\begin{aligned} \left\|\mathbf{y}_i - e^{j\theta_{i,1}}\mathbf{x}_1 - e^{j\theta_{i,2}}\mathbf{x}_2\right\|^2 \approx \|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 + 2|\mathbf{x}_1\mathbf{x}_2^\dagger|\cos\left(\theta_{i,1} - \theta_{i,2} + \angle\mathbf{x}_1\mathbf{x}_2^\dagger\right) \\ - 2|\mathbf{x}_1^\dagger\mathbf{y}_i|\cos\left(\theta_{i,1} - \angle\mathbf{x}_1^\dagger\mathbf{y}_i\right) - 2|\mathbf{x}_2^\dagger\mathbf{y}_i|\cos\left(\theta_{i,2} - \angle\mathbf{x}_2^\dagger\mathbf{y}_i\right). \end{aligned} \tag{4.44}$$

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*[139].

$$\begin{aligned} \mathrm{q}\left(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{y}_i\right) \propto \exp\left(-\tfrac{\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2}{\mathrm{N}_0}\right) \int_{\theta_{i,1}} \exp\left(\frac{2}{\mathrm{N}_0}\left|\mathbf{x}_1^\dagger\mathbf{y}_i\right|\cos\left(\theta_{i,1} - \angle\mathbf{x}_1^\dagger\mathbf{y}_i\right)\right) \\ \int_{\theta_{i,2}} \exp\left(\frac{2}{\mathrm{N}_0}\left|\mathbf{x}_2^\dagger\mathbf{y}_i\right|\cos\left(\theta_{i,2} - \angle\mathbf{x}_2^\dagger\mathbf{y}_i\right)\right) \mathrm{d}\theta_{i,2}\,\mathrm{d}\theta_{i,1}. \end{aligned} \tag{4.45}$$

Proposition 4.3. The likelihood function is shown to be

$$\mathrm{q}\left(\{\mathbf{x}_1,\mathbf{x}_2\},\mathbf{y}\right) \propto \prod_{i=0}^{\mathsf{N_R}-1} \exp\left(-\frac{\|\mathbf{x}_1\|^2+\|\mathbf{x}_2\|^2}{\mathrm{N}_0}\right) \times \mathrm{I}_0\left(\frac{2}{\mathrm{N}_0}|\mathbf{x}_1^{\dagger}\mathbf{y}_i|\right) \times \mathrm{I}_0\left(\frac{2}{\mathrm{N}_0}|\mathbf{x}_2^{\dagger}\mathbf{y}_i|\right). \tag{4.46}$$

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

$$\mathrm{q}\left(e_j(\{\mathbf{x}_1,\mathbf{x}_2\})=b,\mathbf{y}_i\right) = \sum_{\{\mathbf{x}_1,\mathbf{x}_2\}\,\in\,\boldsymbol{\chi}_b^j} \mathrm{q}\left(\{\mathbf{x}_1,\mathbf{x}_2\},\mathbf{y}_i\right). \tag{4.47}$$

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

$$\Lambda^j\left(\mathbf{y}_i\right) = \log\frac{\mathrm{q}\left(e_j(\{\mathbf{x}_1,\mathbf{x}_2\})=0,\mathbf{y}_i\right)}{\mathrm{q}\left(e_j(\{\mathbf{x}_1,\mathbf{x}_2\})=1,\mathbf{y}_i\right)}. \tag{4.48}$$

Corollary 4.2. For simpler processing, the *max-log approximation* of (4.48) is provided in (4.49).

$$\Lambda^j\left(\mathbf{y}\right) = \sum_{b\in\{0,1\}} (-1)^b \left[\max_{\{\mathbf{x}_1,\mathbf{x}_2\}\,\in\,\boldsymbol{\chi}_b^j}\left(\sum_{t=1}^{\mathsf{N_T}=2}\left(\sum_{i=1}^{\mathsf{N_R}}\frac{2}{\mathrm{N}_0}\left|N_p\hat{\mathrm{h}}_{i,t}^{\mathrm{LS}} + \mathbf{x}_t^{(\mathrm{d})\,\dagger}\mathbf{y}_i^{(\mathrm{d})}\right|\right) - \frac{\|\mathbf{x}_t\|^2}{\mathrm{N}_0}\right)\right]. \tag{4.49}$$

Remark 4.3. 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 presented in (3.38) for the general non-coherent fading channel by setting the relative magnitude of the LOS component $\alpha=1$, corresponding to the pure LOS scenario. Furthermore, if the orthogonality between $\mathbf{x}_1$ and $\mathbf{x}_2$ is not assumed, establishing a metric for such a LOS channel when $\theta_{t,r}$ is unknown would be quite challenging.

1.3 Extensions to ($\mathsf{N_T}\times\mathsf{N_R}$) MU-MIMO

Each user, indexed by $u=1,2,\ldots,\mathsf{U}$, is equipped with $\mathsf{N_T}$ antennas, while the receiver has $\mathsf{N_R}$ antennas. In this configuration, the channel matrix of the $u-th$ user is represented by $\mathbf{H}_u\in\mathbb{C}^{\mathsf{N_T^u}\times\mathsf{N_R}}$ whose $(t,r)-th$ element is a fading coefficient between the $i-th$ transmitting antenna and the $j-th$ receiving antenna of the $u-th$ user. In addition, only the channel distribution is known to the receiver. In the same spirit as the previous proposals for pilot-assisted transmission, it is possible to ensure that advanced estimation methods can be used to discover channel state information for each user, either separately or jointly. The multi-user system can be modelled such that the relationship between the signals received and transmitted by the different users is such that the associated block transmission model is:

$$\mathbf{Y} = \sum_{u=1}^{\mathsf{U}} \mathbf{X}_u\mathbf{H}_u + \mathbf{Z}, \quad u=1,2,\ldots,\mathsf{U}. \tag{4.50}$$

In accordance with the joint estimation/detection technique we proposed earlier. However, there are other techniques in the literature which have demonstrated their merits, namely iterative estimation/detection techniques, or the joinly based three search technique with pilotless transmission, or non-coherent detection which does not perform explicit channel estimation, but exploits statistical knowledge of the channel.
In Figure 4.3, we depict the MIMO MAC channel accommodating K users (UEs) or transmitters. The effective performance of these techniques depends on channel conditions, system design, the algorithms used and other environmental factors. The use of advanced estimation-detection techniques in conjunction may be beneficial for optimizing MU-MIMO performance.
User detection in a MU-MIMO system over a fading channel with multiple users can be achieved by maximum likelihood detection. Instead of using the well-known *linear detection* such as zero forcing or the *minimum mean square error* (MMSE) method, we can use the proposed maximum likelihood metric for JED. Indeed, ML detection seeks to maximise the likelihood of the observed data given the channel model and the symbols transmitted for each user. However, the complexity of this method increases exponentially with the number of users and the size of the system.

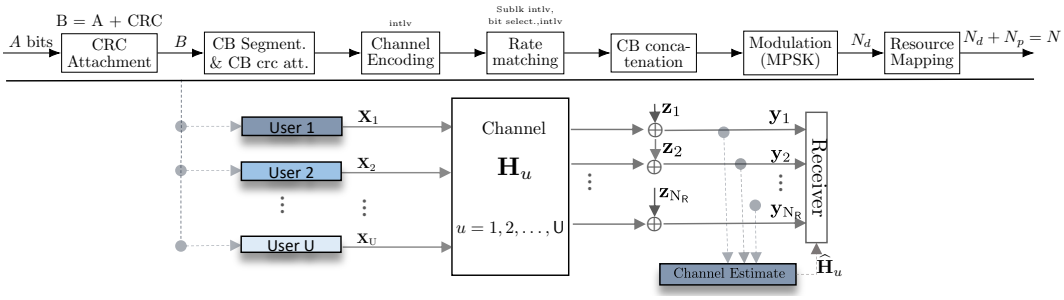


Figure 4.3: Conceptual Illustration of $K-$ users within: $(\mathsf{N_R} \times \mathsf{N_T})$ MAC MIMO system in a block fading environment

Phrased directly, the detection rule for finding $\widehat{\mathbf{X}}_u$ that maximizes the likelihood function can be expressed as follows

$$\widehat{\mathbf{X}}_u = \underset{\mathbf{X}_u}{\operatorname{argmax}}\ \mathsf{q}(\mathbf{X}_u, \mathbf{Y}), \tag{4.51}$$

The derivation of the metric is identical to what we proposed for the SU-MIMO or point-to-point MIMO scenario. However, it should be noted that the transmission here involves multiple users simultaneously, making the communication a *multiple access channel* (MAC) type. Thus, at the reception, the traffic from different users must be detected and decoded separately or jointly. The likelihood function for each user is therefore defined as follows:

$$\mathrm{q}(\mathbf{X}_u, \mathbf{Y}) = \frac{1}{\mathbf{L}_\mathsf{X}} \exp\left(\frac{1}{\mathrm{N}_0}\,\mathrm{tr}\left\{\left(\mathbf{C}_u\widehat{\mathbf{H}}_u^{\mathrm{LS}} + \mathbf{X}_u^{(\mathsf{d})\dagger}\mathbf{Y}^{(\mathsf{d})}\right)^\dagger \mathbf{D}_u \left(\mathbf{C}_u\widehat{\mathbf{H}}_u^{\mathrm{LS}} + \mathbf{X}_u^{(\mathsf{d})\dagger}\mathbf{Y}^{(\mathsf{d})}\right)\right\}\right), \tag{4.52}$$

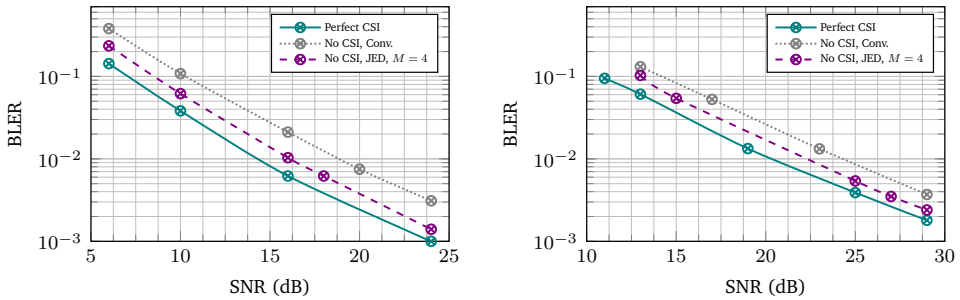
where $\mathbf{D}_u = \left[\mathrm{N}_0\mathbf{I} + \mathbf{X}_u^\dagger\mathbf{X}_u\right]^{-1}$.

1.4 Numerical Results

Figure 4.4a illustrates the performance over a (4×4) MIMO Rayleigh block fading channel under 5G polar-coded modulation. The JED-based receiver achieves a performance gain of approximately 2.75 dB compared to the conventional receiver at a BLER threshold of 0.1%. However, there remains a performance gap of about 1.75 dB between the JED receiver ($M = 4$) and the ideal or *Perfect CSI* receiver

Similarly, Figure 4.4b shows the performance under 5G LDPC-coded modulation. The JED-based receiver provides a gain of approximately 2 dB over the conventional receiver, and the gap between the JED receiver ($M = 4$) and the perfect CSI receiver is around 1.75 dB.

Therefore, it can be also remarkably asserted that the advanced JED-based receiver outperforms the conventional counterpart and demonstrates greater resilience under imperfect channel estimation.



(a) MIMO Rayleigh Block Fading channel, 5G Polar BICM. (b) MIMO Rayleigh block Fading channel, 5G LDPC BICM.

Figure 4.4: 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), (4×4) MIMO.

In addition, considering the JED metric proposed in (4.32) applied to the MIMO configuration, the average time complexity curves shown in Figure 4.5 indicate that the complexity of the JED-based receiver remains noticeably higher than that of the conventional receiver, even when stream-wise processing is employed. It may somehow be important to reduce spatial diversity by reducing the number of transmit antennas (e.g. at most $\mathrm{N}_\top = 2$) to achieve a

good trade-off in terms of performance and complexity within the proposed JED MIMO receiver.

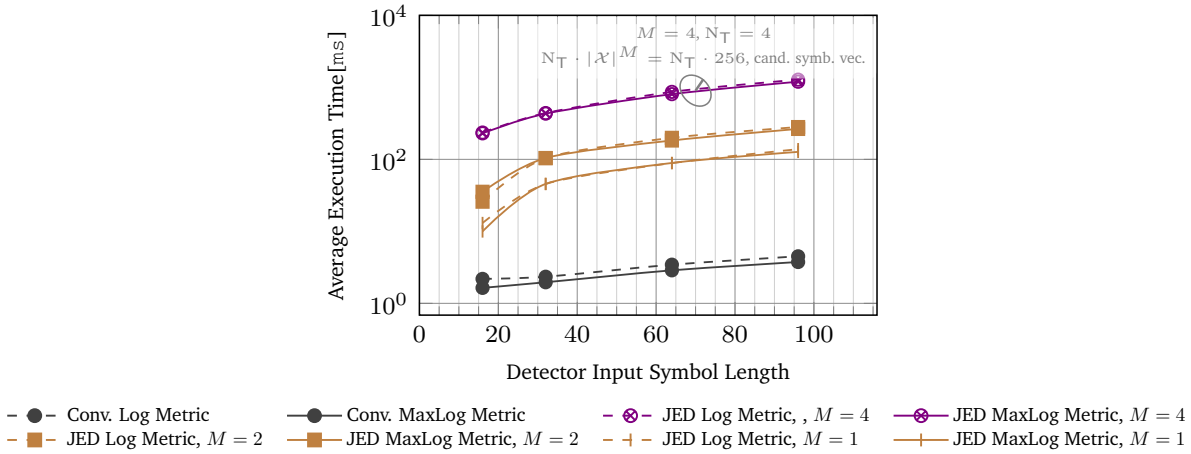


Figure 4.5: Average Time Complexity of the proposed MIMO JED metrics vs Conventional metrics.

2 Summary

This chapter presented BICM metrics tailored for joint estimation and detection in short block-length channels across various scenarios, including SU-MIMO Rayleigh block-fading channels and SU-MIMO Line-of-Sight channels. Addressing cases with unknown channel state information and low pilot density, the performance analysis relied on complete 5G transmitter and receiver chains for both Polar- and LDPC-coded transmissions combined with a QPSK modulation scheme. As described in Chapter 3, the proposed BICM MIMO receiver metrics also employ block-wise joint estimation and detection strategies. The considered transmissions also involved reference signals interleaved with data and sent over a small number of OFDM symbols, making near-perfect channel estimation unattainable. The results demonstrated that, with detection windows spanning approximately four modulated symbols, the proposed BICM MIMO metrics achieved detection performance close to that of a coherent receiver with perfect channel state information for both Polar- and LDPC-coded configurations. Additionally, a comparative analysis of the complexity of the proposed receivers versus conventional receivers in the short-block regime was provided. Overall, the presented joint estimation–detection strategy, built upon an enhanced likelihood metric, offers a significant advancement in receiver design for B5G/6G systems, especially in highly dynamic or resource-constrained environments. The next chapter addresses the design of receivers for short data transmissions using Reed–Muller codes, taking into account both reliability and complexity aspects.

Chapter 5

Receiver Design (Part III)

Contents

1 Maximum Likelihood Receiver

1.1 Non-coherent energy term role

To demonstrate the consequences of neglecting 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[134]. In the transmission process on short block channels, messages with a length of K ranging from 3 to 11 bits pass through a RM encoder characterized by $\mathscr{C}(32, K)$. The transmission framework follows a straightforward procedure and can be formulated as follows:

$$c_\ell = \left(\sum_{k=0}^{K-1} b_k \cdot m_{\ell,k} \right) \bmod 2, \tag{5.1}$$

where $\ell = 0,\ 1,\ \ldots,\ N' - 1$, N'= 32 and $m_{\ell,k}$ represents the basis sequences as defined in [132, Table 5.3.3.3-1]. 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 bit-width 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

i	$M_{i,0}$	$M_{i,1}$	$M_{i,2}$	$M_{i,3}$	$M_{i,4}$	$M_{i,5}$	$M_{i,6}$	$M_{i,7}$	$M_{i,8}$	$M_{i,9}$	$M_{i,10}$
0	1	1	0	0	0	0	0	0	0	0	1
1	1	1	1	0	0	0	0	0	0	1	1
2	1	0	0	1	0	0	1	0	1	1	1
3	1	0	1	1	0	0	0	0	1	0	1
4	1	1	1	1	0	0	0	1	0	0	1
5	1	1	0	0	1	0	1	1	1	0	1
6	1	0	1	0	1	0	1	0	1	1	1
7	1	0	0	1	1	0	0	1	1	0	1
8	1	1	0	1	1	0	0	1	0	1	1
9	1	0	1	1	1	0	1	0	0	1	1
10	1	0	1	0	0	1	1	1	0	1	1
11	1	1	1	0	0	1	1	0	1	0	1
12	1	0	0	1	0	1	0	1	1	1	1
13	1	1	0	1	0	1	0	1	0	1	1
14	1	0	0	0	1	1	0	1	0	0	1
15	1	1	0	0	1	1	1	1	0	1	1
16	1	1	1	0	1	1	1	0	0	1	0
17	1	0	0	1	1	1	0	0	1	0	0
18	1	1	0	1	1	1	1	1	0	0	0
19	1	0	0	0	0	1	1	0	0	0	0
20	1	0	1	0	0	0	1	0	0	0	1
21	1	1	0	1	0	0	0	0	0	1	1
22	1	0	0	0	1	0	0	1	1	0	1
23	1	1	1	0	1	0	0	0	1	1	1
24	1	1	1	1	1	0	1	1	1	1	0
25	1	1	0	0	0	1	1	1	0	0	1
26	1	0	1	1	0	1	0	0	1	1	0
27	1	1	1	1	0	1	0	1	1	1	0
28	1	0	1	0	1	1	1	0	1	0	0
29	1	0	1	1	1	1	1	1	1	0	0
30	1	1	1	1	1	1	1	1	1	1	1
31	1	0	0	0	0	0	0	0	0	0	0

Table 5.1: Basis sequences for $(32, K)$ codes [132, Table 5.3.3.3-1].

is simply repetition. The encoded 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 QPSK modulation, resulting in a block of complex-valued modulation symbols $\mathrm{x}(0),\ \mathrm{x}(1),\ \ldots,\ \mathrm{x}\,(E/2-1)$. 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.

On reception, the messages are recovered using maximum likelihood decoding, which can be computationally demanding.

We describe the detection metric for LOS channel with an unknown phase such as $\mathbf{h}_i = e^{j\theta_i}\mathbf{I}$, where θ_i are assumed to be i.i.d. uniform random variables on $[0, 2\pi)$

$$\mathrm{q}\left(\mathbf{x}, \mathbf{y}_i\right) = p\left(\mathbf{y}_i \mid \mathbf{x}\right) = p\left(\mathbf{y}_i \mid \mathbf{x}, \theta_i\right) p\left(\theta_i\right), \tag{5.2}$$

where we have $p(\theta_i) = \frac{1}{2\pi}$, is a uniform distribution on $[0, 2\pi)$.

The likelihood function is equivalent to

$$\mathrm{q}\left(\mathbf{x}, \mathbf{y}\right) = \prod_{i=0}^{\mathrm{N_R}-1} \frac{1}{2\pi} \int_{\theta_i=0}^{2\pi} \exp\left(-\frac{||\mathbf{y}_i - \mathbf{h}_i\mathbf{x}||^2}{\mathrm{N}_0}\right) \mathrm{d}\theta_i. \tag{5.3}$$

Moreover $||\mathbf{y}_i - \mathbf{h}_i\mathbf{x}||^2 = ||\mathbf{y}_i||^2 + ||\mathbf{h}_i\mathbf{x}||^2 - 2\mathrm{Re}\left(\mathbf{y}_i\mathbf{h}_i^*\mathbf{x}^*\right)$ and considering that $\mathbf{x}^\dagger\mathbf{y}_i = \left|\mathbf{x}^\dagger\mathbf{y}_i\right| e^{j\phi_i}$ where $\phi_i = \angle\mathbf{x}^\dagger\mathbf{y}_i$, the likelihood function after excluding multiplicative terms unrelated to $\mathbf{x}$ can be expressed as follows:

$$\mathrm{q}\left(\mathbf{x}, \mathbf{y}\right) \propto \prod_{i=0}^{\mathrm{N_R}-1} \exp\left(-\frac{\|\mathbf{x}\|^2}{\mathrm{N}_0}\right) \cdot \frac{1}{2\pi} \int_0^{2\pi} \exp\left(\frac{2}{\mathrm{N}_0}\left|\mathbf{x}^\dagger\mathbf{y}_i\right| \cos\left(\phi_i - \theta_i\right)\right) \mathrm{d}\theta_i, \tag{5.4}$$

where $\frac{1}{\pi}\int_{\lambda=0}^{\pi} \exp(zcos(\lambda))\mathrm{d}\lambda = \mathrm{I}_0(\mathrm{z})$, $\mathrm{I}_0(\cdot)$ is the modified Bessel function of the first kind and † denote the complex

conjugate transpose or Hermitian.

$$\mathrm{q}\left(\mathbf{x},\mathbf{y}\right)=\prod_{i=0}^{\mathsf{N_R}-1}\exp\left(-\frac{\left\|\mathbf{x}\right\|^2}{\mathrm{N}_0}\right)\mathrm{I}_0\left(\frac{2}{\mathrm{N}_0}\left|\mathbf{x}^\dagger\mathbf{y}_i\right|\right). \tag{5.5}$$

We typically simplify the likelihood function from (5.5) via a *max-log approximation*. The approximation is obtained by assuming that $\log\left\{\sum_i \exp\left(\lambda_i\right)\right\}\sim\max_i\left\{\lambda i\right\}$, taking into account the exponential approximation of the modified Bessel function of the first kind, I_0.

Theorem 5.1. (Equivalence of linear and quadratic metrics under collinearity)
Let $\mathbf{y}_i\in\mathbb{C}^N$ for $i=0,\ldots,\mathsf{N_R}-1$, and define

$$f(\mathbf{x})=\sum_{i=0}^{\mathsf{N_R}-1}\left|\mathbf{x}^\dagger\mathbf{y}_i\right|,\qquad g(\mathbf{x})=\sum_{i=0}^{\mathsf{N_R}-1}\left|\mathbf{x}^\dagger\mathbf{y}_i\right|^2, \tag{5.6}$$

where the optimization variable $\mathbf{x}$ is constrained to a set $\mathcal{X}\subset\mathbb{C}^N$ satisfying $\|\mathbf{x}\|=$ constant for all $\mathbf{x}\in\mathcal{X}$ (e.g., unit-norm vectors or any constant-modulus constellation).
If all receive vectors are collinear, i.e., there exist $\mathbf{v}\neq\mathbf{0}$ and complex scalars c_i such that $\mathbf{y}_i=c_i\mathbf{v}$ for all i, then the maximizers of f and g coincide:

$$\arg\max_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}) \;=\; \arg\max_{\mathbf{x}\in\mathcal{X}} g(\mathbf{x}). \tag{5.7}$$

Proof. Assume $\mathbf{y}_i=c_i\mathbf{v}$ with $\mathbf{v}\neq\mathbf{0}$. For any $\mathbf{x}\in\mathcal{X}$,

$$f(\mathbf{x})=\sum_i\left|\mathbf{x}^\dagger(c_i\mathbf{v})\right|=\left|\mathbf{x}^\dagger\mathbf{v}\right|\sum_i|c_i|, \tag{5.8}$$

and

$$g(\mathbf{x})=\sum_i\left|\mathbf{x}^\dagger(c_i\mathbf{v})\right|^2=\left|\mathbf{x}^\dagger\mathbf{v}\right|^2\sum_i|c_i|^2. \tag{5.9}$$

Since $\sum_i|c_i|$ and $\sum_i|c_i|^2$ are positive constants independent of $\mathbf{x}$, maximizing either $f(\mathbf{x})$ or $g(\mathbf{x})$ is equivalent to maximizing $|\mathbf{x}^\dagger\mathbf{v}|$ (or its square). The functions $|\mathbf{x}^\dagger\mathbf{v}|$ and $|\mathbf{x}^\dagger\mathbf{v}|^2$ are strictly monotone transforms of one another over nonnegative arguments, and thus have identical maximizers on $\mathcal{X}$. Therefore,

$$\arg\max_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})=\arg\max_{\mathbf{x}\in\mathcal{X}} g(\mathbf{x}). \tag{5.10}$$

■

Corollary 5.1. (Application to Constant-Modulus Constellations.)
If $\mathcal{X}$ is a constant-modulus constellation (e.g., BPSK, QPSK, or any unit-norm beamforming codebook), then the ML metric

$$\hat{\mathbf{x}}=\arg\max_{\mathbf{x}\in\mathcal{X}}\left(\sum_{i=0}^{\mathsf{N_R}-1}\frac{2}{\mathrm{N}_0}\left|\mathbf{x}^\dagger\mathbf{y}_i\right|-\frac{\left\|\mathbf{x}\right\|^2}{\mathrm{N}_0}\right) \tag{5.11}$$

reduces, up to an additive constant, to

$$\hat{\mathbf{x}}=\arg\max_{\mathbf{x}\in\mathcal{X}}\sum_{i=0}^{\mathsf{N_R}-1}\left|\mathbf{x}^\dagger\mathbf{y}_i\right|. \tag{5.12}$$

Under the condition of collinearity on the observations $\mathbf{y}_i$ and applying Theorem 5.1, this metric is also equivalent to

$$\hat{\mathbf{x}}=\arg\max_{\mathbf{x}\in\mathcal{X}}\sum_{i=0}^{\mathsf{N_R}-1}\left|\mathbf{x}^\dagger\mathbf{y}_i\right|=\arg\max_{\mathbf{x}\in\mathcal{X}}\sum_{i=0}^{\mathsf{N_R}-1}\left|\mathbf{x}^\dagger\mathbf{y}_i\right|^2. \tag{5.13}$$

Remark 5.1. If the observations $\mathbf{y}_i$ are not collinear, the maximizers of f and g generally differ. In practice, the $\mathbf{y}_i$ are rarely perfectly collinear (i.e., they are quasi-collinear), but the dominant direction remains that of $\mathbf{x}$, assuming the additive noise is isotropic and does not introduce any preferred direction. Under this condition, both decision metrics yield the same ML decision for constant-modulus constellations (e.g., BPSK, QPSK), which establishes the equivalence of the two metrics in terms of decision.

The metric is a frequency interleaving between data symbols and DMRS symbols. So, if we posit $\mathbf{x} = \mathbf{x}^{(\mathsf{p})} + \mathbf{x}^{(\mathsf{d})}$, then

$$\left|\left(\mathbf{x}^{(\mathsf{d})^\dagger} + \mathbf{x}^{(\mathsf{p})^\dagger}\right)\mathbf{y}_i\right|^2 = \left|\mathbf{x}^{(\mathsf{p})^\dagger}\mathbf{y}_i^{(\mathsf{p})} + \mathbf{x}^{(\mathsf{d})^\dagger}\mathbf{y}_i^{(\mathsf{d})}\right|^2 \tag{5.14}$$

More generally, in the traditional literature, the detection of signals with training information can be extended using (5.14) which is often referred to as an *estimator-correlator*.
The equation encompasses both a *quasi-coherent detection* term and a *non-coherent energy* term.

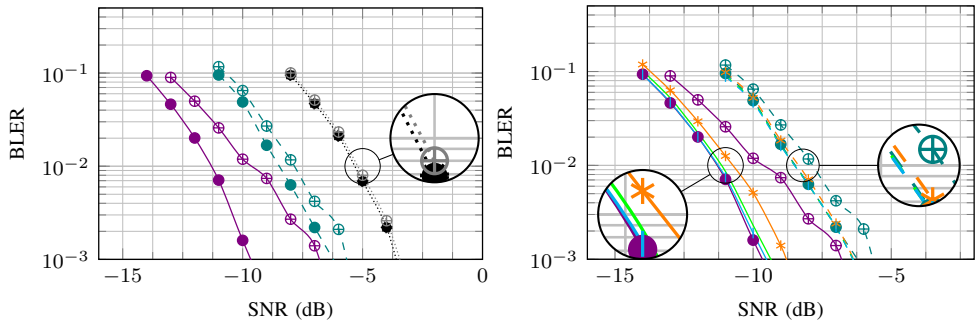
Remark 5.2. Since this quadric expression (5.14) is computationally hard to accomplish, especially when dealing with longer block lengths, even in the short block length regime, the non-coherent data term, as outlined below, is usually neglected. It's worth noting, however, that taking the non-coherent term into account increases receiver sensitivity, so the performance of the underlying receiver will be better.

$$\hat{\mathbf{x}}^{(\mathsf{d})} = \underset{\mathbf{x}^{(\mathsf{d})}}{\operatorname{argmax}} \underbrace{\sum_{i=0}^{\mathsf{N_R}-1}\left|\mathbf{x}^{(\mathsf{p})^\dagger}\mathbf{y}_i^{(\mathsf{p})}\right|^2}_{\text{data-independent term}} + \underbrace{\sum_{i=0}^{\mathsf{N_R}-1}\left|\mathbf{x}^{(\mathsf{d})^\dagger}\mathbf{y}_i^{(\mathsf{d})}\right|^2}_{\text{non-coherent data term}} + \underbrace{2\,\mathrm{Re}\left(\sum_{i=0}^{\mathsf{N_R}-1}\mathbf{x}^{(\mathsf{p})^\dagger}\mathbf{y}_i^{(\mathsf{p})}\cdot\mathbf{y}_i^{(\mathsf{d})\dagger}\mathbf{x}^{(\mathsf{d})}\right)}_{\text{quasi-coherent term}}, \tag{5.15}$$

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

Remark 5.3. Since this quadric is computationally hard to accomplish, especially when dealing with longer block lengths, even in the short block length regime, the non-coherent data term, as outlined below, is usually neglected. It's worth noting, however, that taking the non-coherent data term into account increases receiver sensitivity, so the performance of the underlying receiver will be better.

In what follows, we demonstrate that neglecting the non-coherent data term in current systems leads to a significant performance penalty in typical operating regimes of 5G/6G systems.
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 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 antenna ports were subjected to independent and identically distributed realizations, with no correlation modeling applied. Indeed, TDL-C is a 3GPP reference channel model, which is characterized by a long delay spread used for non-MIMO assessments[135].
The comparison of performances between the optimal ML receiver, which is a fully non-coherent detector, and the conventional receiver, which is a quasi-coherent detector, is presented in Figure 5.1a. The performance difference arises from the non-inclusion of the *non-coherent energy term* (NCET) in the quasi-coherent detector.



(a) Performance Losses w.r.t the non-coherent energy term.

(b) Adaptive DMRS power Adjustement via β

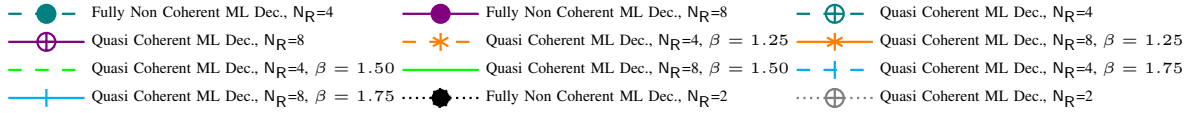


Figure 5.1: Block Error Rate, 4 bits, (SIMO, 3GPP TDL-C urban NLOS channel, 300ns of Delay spread, Unknown *Channel State Information*

In Figure 5.1a, 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 shows that operating in regions of low spectral efficiency while ignoring the non-coherent term leads to significant losses in performance and a reduction in receiver sensitivity. These regions are marked by a low SNR, which significantly alters the accurateness of channel estimates in each reception chain.
Consequently, the significance of the *non-coherent energy term* becomes evident. From this insight, the quasi-coherent receiver is arguably suboptimal in terms of performance, despite its reduced complexity.
However, this observed performance decline could potentially be rectified by enhancing the channel estimate accuracy.

1.2 Performance loss mitigation

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. This is applicable to both downlink and uplink transmissions without incurring any penalty in terms of peak power increase. As illustrated in Figure 5.1b, performance improvement can be observed across various values of β. The optimal performance enhancement is achieved when β is set to 1.75. As indicated by the simulation, when $\mathsf{N_R}$=4 and β is set to 1.5, the suboptimal ML receiver exhibits performance comparable to that of the optimal ML receiver, effectively bridging the initial 0.6 dB gap at the 1% BLER threshold.

1.3 Exhaustive Maximum Likelihood Decoding

At message bit level recovery via ML, it is not necessary to proceed a priori with the detection of symbols $\{\hat{\mathbf{x}}^{(\mathsf{d})}$, we can directly estimate the binary message $\hat{\boldsymbol{m}}$. We perform an exhaustive comparison on the $|\mathcal{M}| = 2^K$ candidate messages. Therefore, the decoder for Reed-Muller codes $\mathscr{C}(32, K)$ uses the following metric to produce a binary message word :

$$\begin{aligned}\hat{\boldsymbol{m}} &= \underset{\boldsymbol{m}\,\in\,\mathcal{M}}{\operatorname{argmax}} \sum_{i=0}^{\mathsf{N_R}-1} \left|\mathbf{x}^{(\mathsf{p})^\dagger}\mathbf{y}_i^{(\mathsf{p})} + \mathbf{x}^{(\mathsf{d})^\dagger}(\boldsymbol{m})\mathbf{y}_i^{(\mathsf{d})}\right|^2, \\ &= \underset{\boldsymbol{m}\,\in\,\mathcal{M}}{\operatorname{argmax}} \sum_{i=0}^{\mathsf{N_R}-1} \left|N_p\hat{\mathrm{h}}_i^{\mathsf{LS}} + \mathbf{x}^{(\mathsf{d})^\dagger}(\boldsymbol{m})\mathbf{y}_i^{(\mathsf{d})}\right|^2\end{aligned} \tag{5.16}$$

The decoding procedure is applied to obtain the most probable message $\hat{\boldsymbol{m}}$ from $\mathcal{M}$ which denotes the set of possible codewords the set of transmitted binary messages.

2 Block-based FoRM Receiver

Reed-Muller (RM) codes are commonly known to be decodable using Hadamard or Fast Hadamard transforms. However, it is well-established that decoding the first-order $\mathbf{RM}(r = 1, \mathsf{M})$ code 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 [110, 147–150].
Although maximum likelihood decoding algorithms have been extensively investigated in traditional literature for decoding data packets encoded with *first-order RM* (ForRM) codes [146], 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* (HT) 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 $\mathbf{RM}(1, \mathsf{M} = 10)$. 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 $\mathbf{RM}(1, \mathsf{M})$ 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.

2.1 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**$(1, \mathsf{M} = 4)$ and **RM**$(1, \mathsf{M} = 5)$, can be employed to encode the respective sub-blocks of 5 bits and 6 bits. In regard to **RM**$(1, 4)$, the codewords are generated using (5.17).

$$\mathbf{c}^{(1)} = \boldsymbol{m}^{(1)}\mathbf{G}^{(1)} = \boldsymbol{m}^{(1)}.\begin{bmatrix} \mathbf{1} & \boldsymbol{v}_4 & \boldsymbol{v}_3 & \boldsymbol{v}_2 & \boldsymbol{v}_1 \end{bmatrix}^{\mathsf{T}}. \tag{5.17}$$

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

Table 5.2: **RM**$(1, \mathsf{M} = 4)$ generator matrix sequences.

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

Thus, the generator matrix is represented as

$$\mathbf{G}^{(1)} = \begin{bmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1 \\ 0&0&0&0&1&1&1&1&0&0&0&0&1&1&1&1 \\ 0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1 \\ 0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1 \end{bmatrix}. \tag{5.18}$$

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

$$\mathbf{c}^{(2)} = \boldsymbol{m}^{(2)}\mathbf{G}^{(2)} = \boldsymbol{m}^{(2)}.\begin{bmatrix} \mathbf{1} & \boldsymbol{v}_5 & \boldsymbol{v}_4 & \boldsymbol{v}_3 & \boldsymbol{v}_2 & \boldsymbol{v}_1 \end{bmatrix}^{\mathsf{T}}. \tag{5.19}$$

Thus, the generator matrix is represented as

$$\mathbf{G}^{(2)} = \begin{bmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1 \\ 0&0&0&0&1&1&1&1&0&0&0&0&1&1&1&1&0&0&0&0&1&1&1&1&0&0&0&0&1&1&1&1 \\ 0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1 \\ 0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1 \end{bmatrix}. \tag{5.20}$$

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

$$\mathbf{c} = [\mathbf{c}^{(1)}\ \mathbf{c}^{(2)}] = [\boldsymbol{m}^{(1)}\mathbf{G}^{(1)}\ \boldsymbol{m}^{(2)}\mathbf{G}^{(2)}]. \tag{5.21}$$

where

$$\begin{aligned} \boldsymbol{m}^{(1)} &= \left(m_0^{(1)}, m_4^{(1)}, m_3^{(1)}, m_2^{(1)}, m_1^{(1)}\right) \xrightarrow{\mathscr{C}(16,5)} \mathbf{c}^{(1)} &&= \left(c_0^{(1)}, c_1^{(1)}, \ldots, c_{15}^{(1)}\right), \\ \boldsymbol{m}^{(2)} &= \left(m_0^{(2)}, m_5^{(2)}, m_4^{(2)}, m_3^{(2)}, m_2^{(2)}, m_1^{(2)}\right) \xrightarrow{\mathscr{C}(32,6)} \mathbf{c}^{(2)} &&= \left(c_0^{(2)}, c_1^{(2)}, \ldots, c_{31}^{(2)}\right). \end{aligned}$$

The procedure of block-based RM encoding is depicted in Figure 5.2.

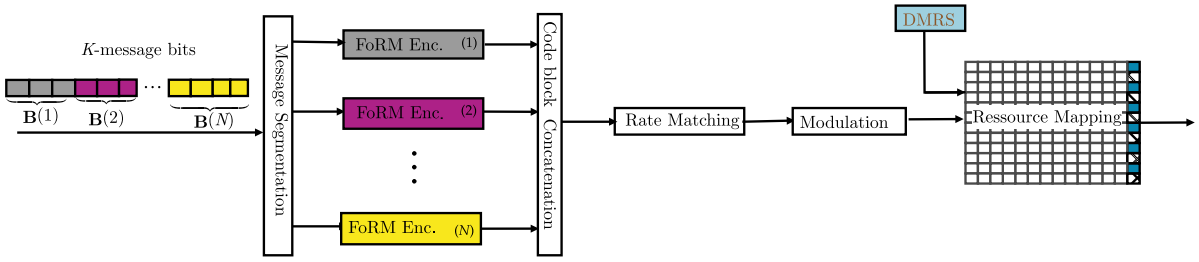


Figure 5.2: Block-based **RM**(1, M) encoding of Short block -length :Transmitter end.

2.2 Block-based Decoding via FHT

A Hadamard matrix, represented as $\mathbf{H}_n$, is a square matrix with dimensions $n \times n$ consisting of elements of ± 1. It satisfies the property that the matrix product of $\mathbf{H}_n$ and its transpose, $\mathbf{H}_n^{\mathsf{T}}$, results in a scalar multiple of the identity matrix, $\mathbf{I}_n$, where $\mathbf{I}_n$ is an $n \times n$ identity matrix and n is the order of the Hadamard matrix. The distinct columns of H are pairwise orthogonal, as are the rows.

$$\mathbf{H}_1 = [1], \quad \mathbf{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad \mathbf{H}_n = \begin{bmatrix} \mathbf{H}_{n-1} & \mathbf{H}_{n-1} \\ \mathbf{H}_{n-1} & -\mathbf{H}_{n-1} \end{bmatrix}, \quad n = 1, 2, 4, 16, 32, \ldots \tag{5.22}$$

The Sylvester construction of Hadamard matrices is a widely recognized mathematical technique with well-established foundations. This construction method facilitates the generation of Hadamard matrices with prescribed dimensions, including notable sizes such as $1, 2, 4, 8, 16, 32$, and so forth. The utilization of this approach leads to the formation of linear codes with inherent linear properties.

Consider the received sequence $\mathbf{u} = (u_0, u_1, \ldots, u_{2^{\mathsf{M}}-1}) \in \mathbb{F}_2$, and let $\mathbf{c} = (c_0, c_1, \ldots, c_{2^{\mathsf{M}}-1}) \in \mathbb{F}_2$ be a codeword. The bipolar representation of $\mathbf{u}$ is denoted as $\mathbf{U} \in \{-1, +1\}$ and is defined as $\mathbf{U} = (-1)^{\mathbf{u}}$. 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} 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 $\mathbf{C}_i$ and constructing the matrix $\mathbf{H} = \begin{bmatrix} \mathbf{C}_0 & \mathbf{C}_1 & \ldots & \mathbf{C}_{2^{\mathsf{M}}-1} \end{bmatrix}$, the computation of all correlations can be expressed as follows:

$$\boldsymbol{\Delta} = \mathbf{U}\mathbf{H}, \tag{5.23}$$

where $\mathbf{H}$ is a matrix of dimension 2^{M}. For the first sub-block, which utilizes a **RM**$(1, 4)$ code with generator matrix $\mathbf{G}^{(1)}$, $\mathbf{H}_{16}^{(1)}$ is employed. Similarly, for the second sub-block, which employs a **RM**$(1, 5)$ code with generator matrix $\mathbf{G}^{(2)}$, $\mathbf{H}_{32}^{(2)}$ is utilized. Furthermore, detailed expositions concerning the algorithms employed in the decoding process of first-order RM codes through the utilization of the Hadamard transform can be readily found within the scholarly works authored by Moon [85] and Wicker [145].

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 $\mathbf{H}_{2^{\mathsf{M}}} = \mathbf{H}_2 \otimes \mathbf{H}_{2^{\mathsf{M}-1}}$, where the Kronecker product of matrices, denoted by $\otimes$, is applied. As a result, the matrix $\mathbf{H}_{2^{\mathsf{M}}}$ can be decomposed as stated in the theorem derived from linear algebraic principles[85].

$$\mathbf{H}_{2^{\mathsf{M}}} = \mathbf{W}_{2^{\mathsf{M}}}^{(1)} \mathbf{W}_{2^{\mathsf{M}}}^{(2)} \cdots \mathbf{W}_{2^{\mathsf{M}}}^{(\mathsf{M})}, \tag{5.24}$$

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

Hence, it follows:

$$\begin{aligned} \mathbf{H}_{16}^{(1)} &= \mathbf{W}_{16}^{(1)} \mathbf{W}_{16}^{(2)} \mathbf{W}_{16}^{(3)} \mathbf{W}_{16}^{(4)} \\ &= \left(\mathbf{I}_{2^3} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^0}\right) \left(\mathbf{I}_{2^2} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^1}\right) \\ &\quad \left(\mathbf{I}_{2^1} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^2}\right) \left(\mathbf{I}_{2^0} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^3}\right). \end{aligned} \tag{5.25}$$

$$\begin{aligned}\mathbf{H}_{32}^{(2)} &= \mathbf{W}_{32}^{(1)}\mathbf{W}_{32}^{(2)}\mathbf{W}_{32}^{(3)}\mathbf{W}_{32}^{(4)}\mathbf{W}_{32}^{(5)} \\ &= \left(\mathbf{I}_{2^4} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^0}\right)\left(\mathbf{I}_{2^3} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^1}\right)\left(\mathbf{I}_{2^2} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^2}\right) \\ &\quad \left(\mathbf{I}_{2^1} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^3}\right)\left(\mathbf{I}_{2^0} \otimes \mathbf{H}_2 \otimes \mathbf{I}_{2^4}\right).\end{aligned} \tag{5.26}$$

Let's consider $\mathbf{U}^{(1)} = [\mathrm{U}_0, \mathrm{U}_1, \ldots, \mathrm{U}_{15}]$ and $\mathbf{U}^{(2)} = [\mathrm{U}_0, \mathrm{U}_1, \ldots, \mathrm{U}_{31}]$, the received sequences to be fed to the decoders. The corresponding Hadamard transforms can then be written as

$$\begin{aligned}\boldsymbol{\Delta}^{(1)} &= \mathbf{U}^{(1)}\mathbf{H}_{16}^{(1)} = \mathbf{U}^{(1)}\left(\mathbf{W}_{16}^{(1)}\mathbf{W}_{16}^{(2)}\mathbf{W}_{16}^{(3)}\mathbf{W}_{16}^{(4)}\right), \\ \boldsymbol{\Delta}^{(2)} &= \mathbf{U}^{(2)}\mathbf{H}_{32}^{(2)} = \mathbf{U}^{(2)}\left(\mathbf{W}_{32}^{(1)}\mathbf{W}_{32}^{(2)}\mathbf{W}_{32}^{(3)}\mathbf{W}_{32}^{(4)}\mathbf{W}_{32}^{(5)}\right),\end{aligned} \tag{5.27}$$

where

$$\begin{aligned}\mathbf{W}_{16}^{(1)} &= \mathbf{I}_8 \otimes \mathbf{H}_2, \\ \mathbf{W}_{16}^{(2)} &= \mathbf{I}_4 \otimes \mathbf{H}_2 \otimes \mathbf{I}_2, \\ \mathbf{W}_{16}^{(3)} &= \mathbf{I}_2 \otimes \mathbf{H}_2 \otimes \mathbf{I}_4, \\ \mathbf{W}_{16}^{(4)} &= \mathbf{H}_2 \otimes \mathbf{I}_8, \\ \mathbf{W}_{32}^{(1)} &= \mathbf{I}_{16} \otimes \mathbf{H}_2 \\ \mathbf{W}_{32}^{(2)} &= \mathbf{I}_8 \otimes \mathbf{H}_2 \otimes \mathbf{I}_2 \\ \mathbf{W}_{32}^{(3)} &= \mathbf{I}_4 \otimes \mathbf{H}_2 \otimes \mathbf{I}_4 \\ \mathbf{W}_{32}^{(4)} &= \mathbf{I}_2 \otimes \mathbf{H}_2 \otimes \mathbf{I}_8 \\ \mathbf{W}_{32}^{(5)} &= \mathbf{H}_2 \otimes \mathbf{I}_{16}.\end{aligned}$$

The Fast Hadamard transform has m stages, so has a computational complexity of $\mathsf{M}2^{\mathsf{M}} = \mathcal{O}\left(N' \log N'\right)$ (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 5.3 and Algorithm 4.

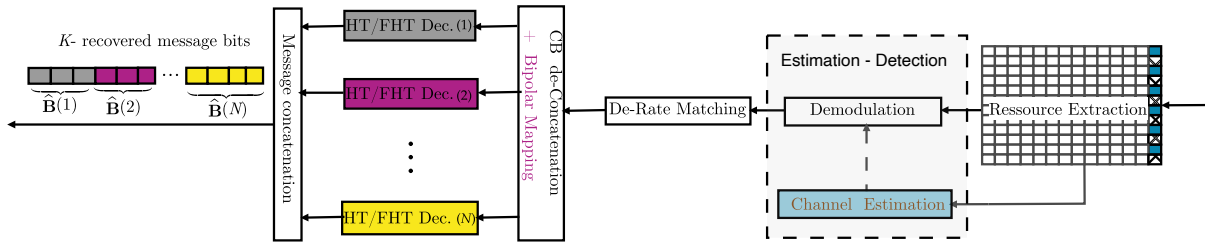


Figure 5.3: Block-based HT/FHT based-decoding of Short blocks : Receiver end.

Algorithm 4: Block based FHT Decoding algorithm for FoRM code $\mathbf{RM}(1, \mathrm{M})$

Require: received code 1: $\mathbf{u}^{(1)} = (u_0, u_1, \ldots, u_{15})$, received code 2: $\mathbf{u}^{(2)} = (u_0, u_1, \ldots, u_{31})$.

Find the bipolar representation of $\mathbf{u}^{(1)}$:$\mathbf{U}^{(1)} = (-1)^{\mathbf{u}^{(1)}}$ and $\mathbf{u}^{(2)}$: $\mathbf{U}^{(2)} = (-1)^{\mathbf{u}^{(2)}}$

Optimize the decoding process via FHT for each sub-decoder.

The first subblock, involves the use of the 16-order Hadamard matrix $\mathbf{H}$.

`Compute FHT:`
$\boldsymbol{\Delta}^{(1)} = \mathbf{U}^{(1)}\mathbf{H}_{16}^{(1)} = \mathbf{U}^{(1)}\left(\mathbf{W}_{16}^{(1)}\mathbf{W}_{16}^{(2)}\mathbf{W}_{16}^{(3)}\mathbf{W}_{16}^{(4)}\right) = (\delta_0^{(1)}, \delta_1^{(1)}, \ldots, \delta_{2^4-1}^{(1)})$

The Second subblock, involves the use of the 32-order Hadamard matrix $\mathbf{H}$.

`Compute FHT:`
$\boldsymbol{\Delta}^{(2)} = \mathbf{U}^{(2)}\mathbf{H}_{32}^{(2)} = \mathbf{U}^{(2)}\left(\mathbf{W}_{32}^{(1)}\mathbf{W}_{32}^{(2)}\mathbf{W}_{32}^{(3)}\mathbf{W}_{32}^{(4)}\mathbf{W}_{32}^{(5)}\right) = (\delta_0^{(2)}, \delta_1^{(2)}, \ldots, \delta_{2^5-1}^{(2)})$

Find the coordinates $\delta_i^{(1)}$ and $\delta_i^{(2)}$ with the largest magnitude.

Let $i^{(1)}$ and $i^{(2)}$ (the indexes of $\delta_i^{(1)}$ and $\delta_i^{(2)}$ resp.) have the binary expansions:

$i^{(1)} = (i_4^{(1)}, i_3^{(1)}, i_2^{(1)} \ldots, i_1^{(1)})_2$ and $i^{(2)} = (i_5^{(2)}, i_4^{(2)}, i_3^{(2)}, i_2^{(2)}, i_1^{(2)})_2$;

if $(\delta_i^{(1)}, \delta_i^{(2)})>0$: (**1** `is not sent`) **then**

$\hat{\mathbf{c}}^{(1)} = \sum_{j=1}^{4} i_j^{(1)} v_j$:= `a linear combinaison of only` v_1, v_2, v_3, v_4 ;

$\hat{\mathbf{c}}^{(2)} = \sum_{j=1}^{5} i_j^{(2)} v_j$:= `a linear combinaison of only` v_1, v_2, v_3, v_4, v_5 ;

else if $(\delta_i^{(1)}, \delta_i^{(2)})<0$: (**1** `is sent-complement all the bits`) **then**

$\hat{\mathbf{c}}^{(1)} = \mathbf{1} + \sum_{j=1}^{4} i_j^{(1)} v_j, \quad \hat{\mathbf{c}}^{(2)} = \mathbf{1} + \sum_{j=1}^{5} i_j^{(2)} v_j$;

return Estimated codeword: $\hat{\mathbf{c}} = [\hat{\mathbf{c}}^{(1)}, \hat{\mathbf{c}}^{(2)}]$.

Furthermore, as we stated earlier, message recovery using maximum likelihood decoding is computationally demanding for both *fully non-coherent* and *quasi-coherent* receivers.

The following simulation results were performed under the assumption of Rayleigh flat fading channel, utilizing 2, 4 and 8 antenna configurations. The antenna ports were subjected to independent and identically distributed realizations, with no correlation modeling applied. Figure 5.4, demonstrate the merits of the proposed block-based RM decoding approach via HT/FHT. These results are contrasted with the conventional RM decoding technique utilizing a ML decoder. The presented results demonstrate that the ML decoder exhibits superior performance. However,

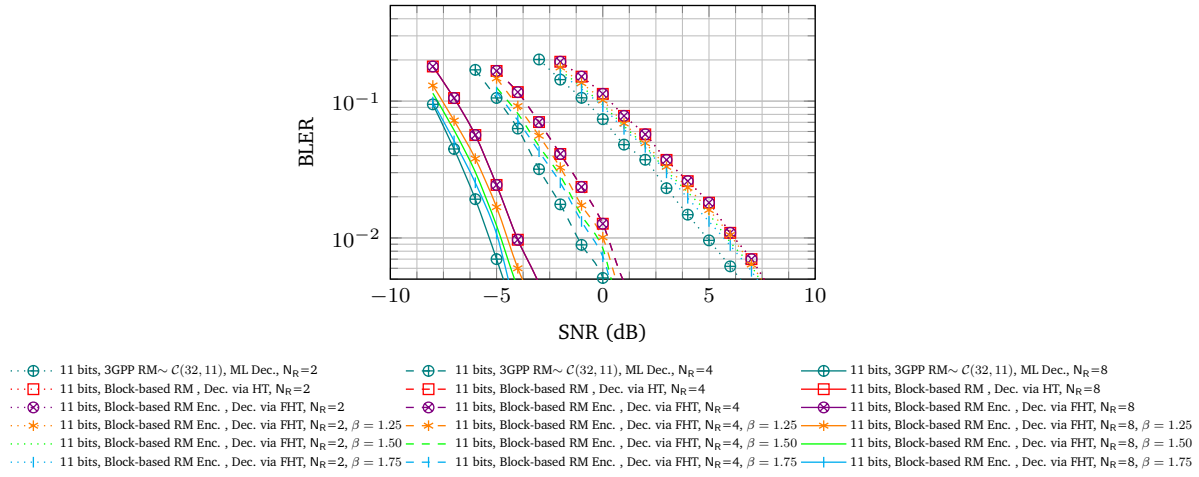


Figure 5.4: Block Error Rate, 11 bits, 2 PRB(16 REs=data, 8 REs=DMRS), Block-based decoding via HT & FHT based decoders vs ML decoder, adaptive DMRS/data power adjustment bia β, ($\{8, 4, 2\} \times 1$) SIMO, Rayleigh fading Channel, Unknown *Channel State Information*.

this advantage is accompanied by a significant increase in computational complexity, even with the block decoding principle. In a practical implementation, each received code block ($\mathscr{C}(48, 11) = [\mathscr{C}_1(16, 5), \mathscr{C}_2(32, 6))]$) is processed by a corresponding decoder, followed by concatenation at the decoder output stage. It is important to note that the

block decoding performance via HT and FHT is comparable, with a difference of approximately 1.3 dB compared to the ML receiver at a BLER threshold of 1% when $\mathsf{N_R} = 4$. Moreover, some of the performance gap with the ML receiver can be bridged or at least reduced.

By combining block-based FHT decoding with adaptive DMRS power adjustment, observable performance improvement is achieved by varying the values of β, particularly $\beta = 1.25, 1.5, 1.75$. The most significant performance gain is observed when $\beta = 1.75$. Simulation has shown that when the number of reception antennas is 4 and $\beta = 1.75$ is selected, the performance of the FHT-based decoder approaches that of the ML decoder at a BLER threshold of 1%, corresponding to an additional gain of 1 dB.

Furthermore, it is important to note that the adaptive DMRS/data power adjustment process yields greater benefits with an increased number of receiving antennas. Consequently, the disparity between the performance of the ML receiver and the FHT-based receiver can be minimized.

Furthermore, in terms of computational efficiency, block decoding using the Fast Hadamard transform is more advantageous as it offers a faster convergence time, which we will endeavor to demonstrate hereafter.

The graph in Figure 5.5 illustrates the number of operations required, denoted as Π, in relation to the input size, denoted as $N' = 2^K$, for both the ML decoder and the FHT-based decoders. Notwithstanding the computational time complexity involved can be assessed through comprehensive numerical simulations within a relatively simplified analytical framework.

This method highlights the number of steps or essential operations required for a decoder to achieve convergence and successfully recover transmitted messages. To facilitate understanding of the graph, we introduce the notations ${N_1}' = 2^4 = 16$ and $N_2' = 2^5 = 32$, representing the number of candidate data words that can be decoded by decoder 1 via FHT and decoder 2 via FHT. These notations correspond to operations $\Pi_1 = (N_1' \log N_1')$ and $\Pi_2 = N_2' \log N_2'$. As for the ML decoder, which is used to recover codewords encoded using the RM 3GPP code family, the case is such that $N' = 2^{11} = 2048$, resulting in $\Pi^{(\mathsf{ML})} = {N'}^2$ operations.

Lastly, focusing on the BFHT decoder and initially assuming that decoders are executed in parallel and, to some extent, independently of each other, as illustrated in Figure 5.3, this strategy is employed to leverage the merit of parallelism in processing, thereby leading to a diminished inherent complexity. For the BFHT decoding with $N' = N_1' + N_2'$, the overall decoding complexity is represented by $\mathcal{O}\left(\max\left(N_1' \log N_1', N_2' \log N_2'\right)\right)$, which is equivalent to $\Pi^{(\mathsf{FHT})} = (N_2' \log N_2')$, representing the complexity of the decoder processing the longest codewords.

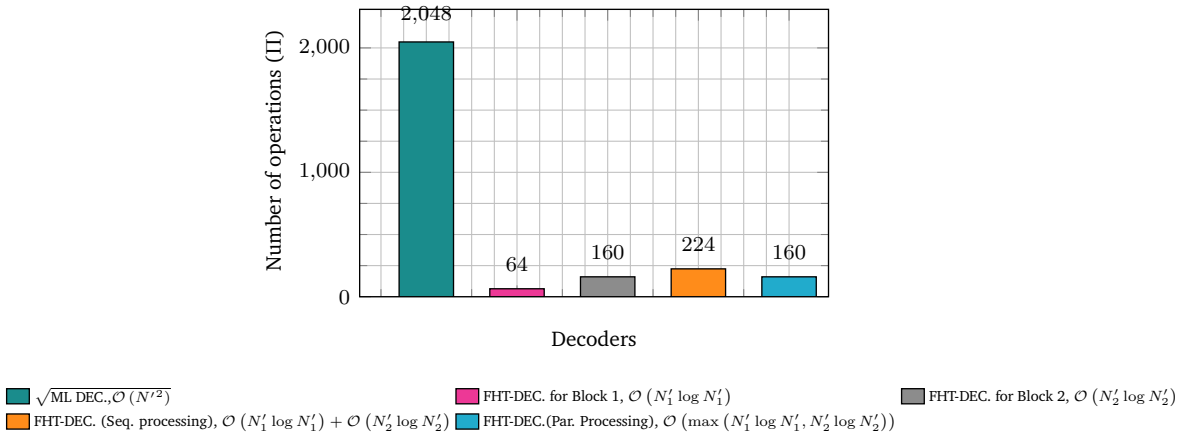


Figure 5.5: Comparative analysis, showing the number of Π operations for the ML decoder compared to block-based FHT-decoders featuring sequential and parallel processing.

The graph illustrates that the ML decoder requires $(2^1 1)^2 = 2048^2$ operations, compared to 160 for the FHT decoder using block decoding. The number of operations required by the ML decoder far exceeds that of the BFHT decoder, even if the underlying decoders are executed sequentially. Therefore, the advantages of FHT decoding are noticeable in terms of complexity.

3 Sparse-aware Vector Coding

Sparse Vector Coding (SVC) is a technique that represents information using a sparse vector $\mathbf{x} \in \mathbb{R}^n$, where only $k \ll n$ entries are nonzero. The SVC techniques are gaining popularity due to their ability to convey information efficiently with reduced bandwidth, low latency, low complexity, and high reliability. Their potential to achieve very low error rates makes them highly suitable for URLLC applications [151–156]. However, conventional SVC methods, as presented in the seminal work of Ji *et al.* [151], often rely on random measurement matrices typically Gaussian or Bernoulli ensembles owing to their theoretical guarantees for sparse recovery. Despite these advantages, random matrices present practical challenges: they require substantial storage and computational resources, and their inherent randomness hinders structured hardware implementation. Furthermore, in finite blocklength regimes, decoding complexity increases and performance tends to degrade. (SVC) techniques, which encode information into sparse vectors, are gaining popularity due to their ability to convey information efficiently with reduced bandwidth, low latency, low complexity, and high reliability. Since its introduction in [151], SVC has been extensively investigated as an efficient transmission paradigm for short-packet URLLC [152–156]. Several extensions have focused on improving reliability by enriching the information-bearing structure of sparse vectors. In particular, enhanced SVC (ESVC) and sparse superposition–based schemes encode information jointly in the support and in the non-zero values, often relying on complex-valued constellations at the support elements to increase spectral efficiency [152, 153]. While these approaches achieve notable BLER gains, their reliance on complex signaling and sophisticated decoding may increase processing latency and hardware complexity, which can be undesirable in ultra-low-latency regimes. In parallel, conventional SVC methods, as presented in the seminal work of Ji *et al.* [151], often rely on random measurement matrices, typically Gaussian or Bernoulli ensembles, owing to their theoretical guarantees for sparse recovery. Despite these advantages, random matrices present practical challenges: they require substantial storage and computational resources, and their inherent randomness hinders structured hardware implementation. Complementary to symbol-domain and code-design enhancements, recent works have highlighted the role of structured measurement matrices in SVC and ESVC to reduce complexity and enable efficient implementations [154]. Structured measurement matrices for SVC, such as partial Hadamard matrices, therefore spark growing interest. Their deterministic construction enables fast transforms and low-complexity implementations. Nonetheless, classical decoding methods based on fixed correlations treat all measurements equally, which can lead to suboptimal performance when noise disproportionately corrupts certain measurements.

Motivated by these observations, the present work focuses on a real-valued, bipolar ESVC framework that is particularly well suited for short-packet transmission under stringent latency constraints. Unlike prior works that primarily enhance SVC through constellation design or code and spreading matrix construction, our contribution targets the receiver stage by introducing an adaptive decoding framework leveraging LLR-based attention-weighted Hadamard projections that dynamically reweight received signal measurements according to their reliability. In this respect, the attention weights are derived from per-element log-likelihood ratios (LLRs) that quantify the reliability of each received signal observations while accounting for the statistics of the transmitted sparsity-aware signal. Simulation results demonstrate that the proposed LLR-based attention-weighted projection decoding consistently outperforms baseline projection decoding for both structured (partial Hadamard) and random (Gaussian) spreading matrices, particularly in challenging transmission scenarios involving frequency-selective fading channels. Moreover, the analysis is extended to a short-packet transmission framework with payloads of 3–11 bits over the 3GPP physical uplink control channel (PUCCH), comparing the proposed approach with the current standard based on 3GPP Reed–Muller coded transmission. We find that SVC using a partial Hadamard spreading matrix combined with LLR-based attention-weighted projection decoding can competitively rival 3GPP RM codes under optimal maximum likelihood (ML) decoding at very low block error rate (BLER) targets. From a computational complexity perspective, the proposed SVC decoder exhibits significant advantages over conventional ML decoding methods, making it a promising ultra-reliable, low-latency short-packet coding solution for next-generation communication systems.

3.1 General Framework of Sparse Vector Coding

A binary message of length B is mapped to a sparse vector $\mathbf{x} \in \mathbb{R}^n$ with sparsity level $k = \|\mathbf{x}\|_0$, which counts its non-zero entries. The support of $\mathbf{x}$, denoted $\mathrm{supp}(\mathbf{x}) = \mathcal{S} \subset \{1, \ldots, n\}$, satisfies $|\mathcal{S}| = k$. Each entry in the support,

$i \in \mathcal{S}$, is assigned a bipolar amplitude $\mathrm{x}_i \in \{-1, +1\}$. The total number of bits per message, including sign bits, is

$$B = \underbrace{\left\lfloor \log_2 \binom{n}{k} \right\rfloor}_{\texttt{Position bits}} + \underbrace{k}_{\texttt{Sign bits}} \quad \text{[bits]}. \tag{5.28}$$

Formally, the binary message is encoded into a sparse vector through the following two-step procedure:

- `Support (active position) selection`: The support set $\mathcal{S}$ is chosen from the $|\mathcal{C}| = \binom{n}{k}$ possible combinations, thereby encoding a first message block $\boldsymbol{m}_1$ of $\lfloor \log_2 \binom{n}{k} \rfloor$ bits.
- `Bipolar modulation (sign assignment)`: Each active position is assigned a symbol ± 1 from a binary antipodal alphabet, thereby encoding a second message block $\boldsymbol{m}_2$ of k bits.

Thus, the total message structure $\boldsymbol{m} \in \mathbb{F}_2^{\lfloor \log_2 \binom{n}{k} \rfloor + k}$ can be expressed as $\boldsymbol{m} = [\boldsymbol{m}_1 \;\; \boldsymbol{m}_2]$ of B bits.
The sparse vector $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\mathrm{x}_i = \begin{cases} +1 & \text{if } i \in \mathcal{S} \text{ and the associated sign bit is } 0, \\ -1 & \text{if } i \in \mathcal{S} \text{ and the associated sign bit is } 1, \\ 0 & \text{otherwise.} \end{cases} \tag{5.29}$$

This mapping establishes a bijection between the binary message $\boldsymbol{m} \in \mathcal{M}$ and the sparse vector $\mathbf{x} \in \mathcal{X}$, since each message $\boldsymbol{m}$ corresponds to exactly one sparse vector $\mathbf{x}$, and vice versa.

$$\mathcal{L} : \mathcal{M} \to \mathcal{X} \quad | \quad \mathbf{x} = \mathcal{L}(\boldsymbol{m}), \tag{5.30}$$

where $\mathcal{M}$ denotes the set of transmitted binary messages, and $\mathcal{X} = \{\mathbf{x} \in \{-1, 0, +1\}^n : \|\mathbf{x}\|_0 = k\}$ represents the set of encoded sparse vectors. The function $\mathcal{L}$ is a discrete mapping implemented as a lookup table (LUT), defining a bijective correspondence between the binary message $\boldsymbol{m}$ and the sparse vector $\mathbf{x}$. SVC, as defined here and conceptually illustrated in Figure 5.6 for parameters $n = 8$ and $k = 2$, can be viewed as a *combinatorial encoding combined with bipolar modulation*.
The mapping between binary message combinations and sparse-vector pair-wise patterns, implementable as an LUT, is given in Table 5.3. Positions marked with ⊠ in the sparse vector indicate the locations of the bipolar symbols, thereby defining the support of the sparse vector.

Table 5.3: SVC mapping table : $n = 8, k = 2$.

$\boldsymbol{m}_1 : \lfloor \log_2 \binom{n}{k} \rfloor$ bits	DEC Idx:$(\boldsymbol{m}_1)_{10}$	Active Pos : $\mathcal{C}(\mathrm{Idx}+1, :)$	SVC Mapping Patterns
0000	0	(1 , 2)	⊠ ⊠ 0 0 0 0 0 0
0001	1	(1 , 3)	⊠ 0 ⊠ 0 0 0 0 0
⋮	⋮	⋮	⋮
0101	5	(1 , 7)	⊠ 0 0 0 0 0 ⊠ 0
⋮	⋮	⋮	⋮
1111	15	(3 , 6)	0 0 ⊠ 0 0 ⊠ 0 0
-	16	(3 , 7)	0 0 ⊠ 0 0 0 ⊠ 0
⋮	⋮	⋮	⋮
-	27	(7 , 8)	0 0 0 0 0 0 ⊠ ⊠

3.2 Structured SVC with Partial Hadamard Matrices

Partial Hadamard matrices are effectively well suited to overcome the limitations inherent in random measurement matrices in SVC. Let $\mathbf{H}_n \in \{\pm 1\}^{n \times n}$ denote the Hadamard matrix of order n, which is a structured orthogonal matrix. The projection matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is constructed from $\mathbf{H}_n$ by randomly selecting $m \leqslant n$ rows of $\mathbf{H}_n$:

$$\mathbf{A} := \frac{1}{\sqrt{m}} \mathbf{P}_m \mathbf{H}_n, \tag{5.31}$$

where $\mathbf{P}_m \in \{0, 1\}^{m \times n}$ is a selection matrix composed of m rows of the identity matrix corresponding to the chosen rows, i.e., $\mathbf{P}_m$ is a partial isometry. The normalization factor $\frac{1}{\sqrt{m}}$ ensures that each column of $\mathbf{A}$ has approximately unit norm (i.e., $\forall i \; \|\mathbf{a}_i\|_2^2 = 1$) and $\forall (i, j)$ the entry $a_{i,j} = \pm \frac{1}{\sqrt{m}}$. The *coherence* of $\mathbf{A}$: $\mu(\mathbf{A}) = \max_{i \neq j} |\mathbf{a}_i^\top \mathbf{a}_j|$ bounds

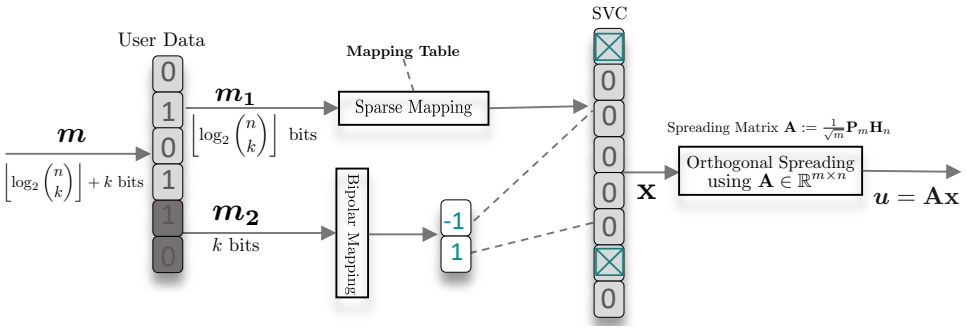


Figure 5.6: Conceptual illustration of Sparse Vector Coding ($n = 8$, $k = 2$).

the cross terms by $|\mathbf{a}_i^\top \mathbf{a}_\ell| \leq \mu$, s.t. $\mu \leq \sqrt{\frac{\log n}{m}}$.
The sparse vector is multiplied by the spreading matrix also referred to as the projection, sensing, or measurement matrix to generate a compressed signal, as in compressive sensing. At the receiver, the original sparse vector is reconstructed through sparse recovery. Conventional decoders are greedy or convex method-based like the well-known orthogonal matching pursuit (OMP) [157], and typically operate under a generic additive white Gaussian noise (AWGN) assumption and do not account for specific channel impairments like fading or interference.
In what follows, we show that SVC, employing orthogonal spreading via a partial Hadamard matrix, can be efficiently decoded using a sophisticated strategy based on LLR-driven attention-weighted projections. The proposed decoder adapts to the statistical characteristics of the transmitted signal and the channel state by utilizing LLRs.

3.3 System modeling and SVC-based Receiver design

The system under consideration encodes a sparse vector $\mathbf{x} \in \{-1, 0, +1\}^n$ with sparsity $k = \|\mathbf{x}\|_0 \ll n$ into a measurement vector $\mathbf{u} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m$ via multiplication by a partial Hadamard matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$. The resulting signal is impaired by a memoryless complex fading channel, where the real and imaginary components of h_i have variance σ_h^2 each, with means given by the real and imaginary parts of μ_h, respectively. The channel is modeled by a diagonal fading matrix $\mathbf{H} = \mathrm{diag}(\mathrm{h}_1, \mathrm{h}_2, \ldots, \mathrm{h}_m)$, where $\mathrm{h}_i \sim \mathbb{CN}(\mu_h, 2\sigma_h^2)$, and is subsequently transmitted over an additive white Gaussian noise channel $\mathbf{z} \sim \mathbb{CN}(0, 2\sigma^2\mathbf{I}_m)$. The received signal can thus be expressed as

$$\mathrm{y}_i = \mathrm{h}_i\,(\mathbf{a}_i^\top \mathbf{x}) + \mathrm{z}_i = \mathrm{h}_i\,u_i + \mathrm{z}_i \quad i = 1, 2, \ldots, m, \tag{5.32}$$

where $\mathbf{a}_i^\top \in \mathbb{R}^{1\times n}$ denotes the i-th row of $\mathbf{A}$ and $u_i = \langle \mathbf{a}_i, \mathbf{x}\rangle = \sum_{j\in\mathcal{S}} a_{ij}x_j$ is a sum of at most k terms, each being $\pm 1/\sqrt{m}$, hence $|u_i| \leq k/\sqrt{m}$.
Equivalently, (5.32) can be expressed in a more compact vectorized form as

$$\mathbf{y} = \mathbf{h} \odot (\mathbf{A}\mathbf{x}) + \mathbf{z} = \mathbf{H}\mathbf{A}\mathbf{x} + \mathbf{z}, \tag{5.33}$$

where $\odot$ denotes the Hadamard (element-wise) product.

At the receiver, the objective is to reliably recover $\mathbf{x}$ from the noisy observations $\mathbf{y}$. A classical approach in sparse vector coding consists of estimating the support of the sparse vector.using the *correlation* or *projection* score computation. This must be achieved with a low-complexity algorithm suitable for short-packet transmission scenarios, where both latency and computational resources are constrained.

3.3.1 Baseline Projection Decoding (Simplified OMP)

First, in the considered scenario, namely transmission over fading channels, an equalization procedure is applied to mitigate channel-induced impairments:

$$\tilde{\mathrm{y}}_i = \Re\left(\frac{\mathrm{h}_i^* \cdot \mathrm{y}_i}{|\mathrm{h}_i|^2 + \varepsilon}\right), \quad \forall\, i, \quad (\mathrm{y}_i, \mathrm{h}_i) \in \mathbb{C}^2, \quad \forall\, \varepsilon > 0. \tag{5.34}$$

The baseline decoding technique, referred to as projection-based support recovery (simplified OMP), involves using the projection matrix $\mathbf{A}$ to correlate with the equalized received signal.

$$\mathbf{c} = \mathbf{A}^\top \tilde{\mathbf{y}} \quad \text{where} \quad \mathrm{c}_j = \mathbf{a}_j^\top \tilde{\mathbf{y}} = \langle \mathbf{a}_j, \tilde{\mathbf{y}}\rangle,\ j = 1, 2, \ldots, n. \tag{5.35}$$

$\mathbf{a}_j \in \mathbb{R}^m$ denotes the $j-th$ column of $\mathbf{A}$, and $\mathbf{c} \in \mathbb{R}^n$ contains the correlation scores.
The k indices with the largest correlation magnitudes are subsequently selected to form the estimated support $\widehat{\mathcal{S}}$.

$$\widehat{\mathcal{S}} = \underset{\mathcal{S}\subset\{1,\ldots,n\},|\mathcal{S}|=k}{\operatorname{argmax}} \sum_{j\in\mathcal{S}=\operatorname{supp}(\mathbf{x})} |\mathrm{c}_j| = \operatorname{top}_k\left(|\mathbf{c}|\right). \tag{5.36}$$

where the operator $\operatorname{top}_k(|\cdot|)$ extracts indices of the k largest-magnitude entries.
Finally, a *naïve* estimate of the sparse vector $\mathbf{x}$ is obtained by assigning signs to the entries corresponding to the estimated support.

$$\hat{\mathrm{x}}_j = \begin{cases} \operatorname{sign}(\mathrm{c}_j) & \text{if } j \in \widehat{\mathcal{S}}, \\ 0 & \text{Otherwise.} \end{cases} \tag{5.37}$$

At the message bit-level recovery stage, LUT-based demapping via the inverse function $\mathcal{L}^{-1}$ can be employed to reconstruct the original binary message. This is made possible by $\mathcal{L}$'s bijective nature, which guarantees a unique inverse mapping:

$$\mathcal{L}^{-1} : \mathcal{X} \to \mathcal{M} \quad | \quad \hat{\boldsymbol{m}} = \mathcal{L}^{-1}(\hat{\mathbf{x}}). \tag{5.38}$$

3.3.2 LLR-Based Attention Weighted Projection Decoding

The conventional support recovery based on naive estimation via baseline projection scores is sensitive to column correlations in $\mathbf{A}$, particularly in the presence of noise, challenging channel conditions, or partial misdetection of the true support (i.e., $\widehat{\mathcal{S}} \neq \mathcal{S}$). To address this limitation, we propose a refined decoding strategy using LLR-driven attention-weighted projections, where adaptive weights derived from per-element LLRs reflect the reliability of each received signal observation. Formally, the approach assigns different weights to the elements of the received vector based on a reliability assessment of each element. In other respects, the LLR computation must account for the actual constellation statistics (i.e., the symbol distribution) of the measurement vector $\mathbf{u}$. Let $u_i = \mathbf{a}_i^\top \mathbf{x}$ be a discrete distribution with the set of possible values $u_i \in \left\{ \frac{2l-k}{\sqrt{m}} \;\middle|\; l = 0, 1, \ldots, k \right\}$ with probability $\mathbb{P}\left(u_i = \frac{2l-k}{\sqrt{m}}\right) = \binom{k}{l}\frac{1}{2^k}$, for all $k \ll m$. Hence, the actual range is $u_i \in [-\boldsymbol{s}, 0, +\boldsymbol{s}]$ where $\boldsymbol{s}$ represents the typical signal levels. Indeed, $\max_{s>0}(\boldsymbol{s}) = \frac{+k}{\sqrt{m}}$ and $\min_{s<0}(\boldsymbol{s}) = \frac{-k}{\sqrt{m}}$.
Formally, we adopt a generalized ternary hypothesis: (1) *Null hypothesis* ($\boldsymbol{\mathcal{H}}_0$): $u_i = 0$, (2) *Positive sign hypothesis* ($\boldsymbol{\mathcal{H}}_+$): $u_i = +s$, and (3) *Negative sign hypothesis* ($\boldsymbol{\mathcal{H}}_-$): $u_i = -s$, with probabilities $\mathbb{P}\left(u_i = \pm s\right) = \frac{1}{2}\left(1 - \mathbb{P}\left(u_i = 0\right)\right)$, $\mathbb{P}\left(u_i = 0\right) = \binom{k}{k/2}\frac{1}{2^k}$, for all even k.
Under the coherent Gaussian model described in equation (5.32), the complex Gaussian probability density function can be expressed as $p(\mathrm{y}_i \mid u_i, \mathrm{h}_i) := \frac{1}{2\pi\sigma^2}\exp\left(-\frac{|\mathrm{y}_i - \mathrm{h}_i u_i|^2}{2\sigma^2}\right)$.
However, note that the LLR $\{\Lambda_i\}_{i=1}^m$ we are seeking is for the sign, even though in reality u_i can be equal to zero. We will therefore consider an LLR that compares the hypotheses $\boldsymbol{\mathcal{H}}_+ : u_i > 0$ and $\boldsymbol{\mathcal{H}}_- : u_i < 0$, ignoring $\boldsymbol{\mathcal{H}}_0 : u_i = 0 \Rightarrow \mathrm{y}_i = \mathrm{z}_i$ since in this case the LLR reduces to pure noise projected onto the direction of h_i. We rather want a metric that reflects the reliability of the sign estimation. In the context of weighting, we want a function that is positive if the estimated sign is positive, negative if negative, and whose absolute value reflects the reliability. The LLR is then obtained by applying Bayes' rule a priori, conditioned on h_i:

$$p(u_i = \pm s \mid \mathrm{y}_i, \mathrm{h}_i) = \frac{p(\mathrm{y}_i \mid u_i = \pm s, \mathrm{h}_i)\, p(u_i = \pm s)}{p(\mathrm{y}_i \mid \mathrm{h}_i)}. \tag{5.39}$$

Taking the ratio and logarithm, we obtain:

$$\Lambda_i = \log \frac{p(\mathrm{y}_i \mid u_i = +s, \mathrm{h}_i)\, p(u_i = +s)}{p(\mathrm{y}_i \mid u_i = -s, \mathrm{h}_i)\, p(u_i = -s)}. \tag{5.40}$$

Assuming equal priors, $p(u_i = +s) = p(u_i = -s)$, they cancel out, yielding:

$$\Lambda_i = \log \frac{p(\mathrm{y}_i \mid u_i = +s, \mathrm{h}_i)}{p(\mathrm{y}_i \mid u_i = -s, \mathrm{h}_i)}. \tag{5.41}$$

Substituting the probability density function into (5.41) and expanding the quadratic terms gives the simplified LLR

expression for the i-th measurement:

$$\begin{aligned}\Lambda_i &= \log\left[\exp\left(-\frac{|\mathrm{y}_i-\mathrm{h}_i(+s)|^2}{2\sigma^2}\right)\right]-\log\left[\exp\left(-\frac{|\mathrm{y}_i-\mathrm{h}_i(-s)|^2}{2\sigma^2}\right)\right]\\ &= \frac{-|\mathrm{y}_i-\mathrm{h}_i(+s)|^2+|\mathrm{y}_i-\mathrm{h}_i(-s)|^2}{2\sigma^2}\\ &= \frac{-|\mathrm{y}_i|^2-s^2|\mathrm{h}_i|^2+2s\Re(\mathrm{h}_i^*\mathrm{y}_i)+|\mathrm{y}_i|^2+s^2|\mathrm{h}_i|^2+2s\Re(\mathrm{h}_i^*\mathrm{y}_i)}{2\sigma^2},\\ &= \frac{2\cdot s\cdot\Re(\mathrm{h}_i^*\mathrm{y}_i)}{\sigma^2},\quad \forall i=1,2,\ldots,m.\end{aligned} \tag{5.42}$$

Remark 5.4. Notwithstanding, the LLR expression (5.42) assumes that the amplitude of u_i is exactly equal to s (i.e., the maximum amplitude). In practice, however, the amplitude may vary, particularly when $k>2$. A more accurate approach is therefore to use the *conditional mean amplitude*, i.e., the *conditional expectation*, which accounts for the statistical distribution of the transmitted signal and better captures its average magnitude.

Proposition 5.1. (LLR Metric for Sparse-Aware Signaling.)
In sparse-aware signaling under the typical sparsity regime $k\ll m$, the LLR reflecting the reliability of the sign estimates for the support of the sparse vector $\mathbf{x}$ is computed by considering the conditional expectation of the magnitude of the transmitted signal observations, namely the measurement entries u_i, as follows:

$$\Lambda_i \approx \frac{2\,\Re\{\mathrm{h}_i^*\mathrm{y}_i\}}{\sigma^2}\cdot\mathbb{E}\big[|\mathbf{a}_i^\top\mathbf{x}|\;\big|\;\mathbf{x}\neq 0\big]=\frac{2\,\Re\{\mathrm{h}_i^*\mathrm{y}_i\}}{\sigma^2}\cdot\mathbb{E}\big[|u_i|\;\big|\;\mathbf{x}\neq 0\big], \tag{5.43}$$

where explicitly $\mathbb{E}[|u_i|\mid\mathbf{x}\neq 0]=\frac{k}{2^{k-1}\sqrt{m}}\binom{k-1}{\lfloor(k-1)/2\rfloor}$.

Indeed, the conditional expectation $\mathbb{E}[|u_i|\mid\mathbf{x}\neq 0]$ arises from the statistical nature of the transmitted signal, capturing its average scale.

Proof. (LLR metric derivation for sparse-aware signaling).

(i) Let's show that $\mathbb{E}[|u_i|\mid\mathbf{x}\neq 0]=\frac{1}{\sqrt{m}}\mathbb{E}[|S_k|]=\frac{2}{2^k\sqrt{m}}\sum_{\ell=\lceil k/2\rceil}^{k}(2\ell-k)\binom{k}{\ell},\quad\forall k\geq 1.$

For a sparse vector $\mathbf{x}$ with support $\mathcal{S}$ of size $k\ll m$: considering the transmitted symbol for measurement i is $u_i=\mathbf{a}_i^\top\mathbf{x}=\sum_{j\in\mathcal{S}}a_{ij}\mathrm{x}_j$, $\forall a_{ij}=\pm 1/\sqrt{m}$ and $\mathrm{x}_j\in\{\pm 1\}$ for $j\in\mathcal{S}$. Hence $u_i=\frac{S_k}{\sqrt{m}}$. By writing $S_k=\sum_{j\in\mathcal{S}}\xi_j$ where $\xi_j=\sqrt{m}\cdot a_{ij}\mathrm{x}_j\sim\pm 1$ (i.i.d. Rademacher variables), then these are i.i.d. uniform random variables on $[-1,+1]$ called Rademacher variables (i.e., following a Rademacher distribution which is a discrete probability distribution where $\mathbb{P}(\xi_j=+1)=1/2$ and $\mathbb{P}(\xi_j=-1)=1/2$). In other words, the sum $S_k=\sum_{j=1}^{k}\xi_j$ is a series (that is, a sum) of Rademacher distributed variables that can be regarded as a simple symmetrical random walk.
Moreover, we reconsider the predefined hypotheses as follows: (1) null hypothesis: $\boldsymbol{\mathcal{H}}_0:\mathrm{x}_j=0$ for all j, (2) positive sign hypothesis: $\boldsymbol{\mathcal{H}}_+:\mathrm{x}_j\neq 0$, sign(u_i) mostly positive and (3) negative sign hypothesis: $\boldsymbol{\mathcal{H}}_-:\mathrm{x}_j\neq 0$, sign(u_i) mostly negative. Indeed, the conditional expectation $\mathbb{E}[|u_i|\mid\mathbf{x}\neq 0]$ should a priori be simply equivalent to $\mathbb{E}[|u_i|]$ because if $\mathbf{x}\neq 0$ then u_i is non-zero. But this is not always the case. Conditionally on $\mathbf{x}\neq 0$, u_i is not always non-zero (there could be cancellations). However, in this context, we are interested in the expectation of the absolute value given that the vector $\mathbf{x}\neq 0$ (i.e., there is at least one non-zero) and that the support is fixed. In reality, we want to condition on the fact that the support is $\mathcal{S}$ (of size k) and that the signs are random. Therefore:

$$\mathbb{E}\left[|u_i|\mid\mathcal{S}\right]=\mathbb{E}\left[\left|\frac{1}{\sqrt{m}}S_k\right|\right]=\frac{1}{\sqrt{m}}\mathbb{E}\left[|S_k|\right].$$

The task is therefore to compute $\mathbb{E}\left[|S_k|\right]$. Indeed, the variable S_k takes even values if k is even and odd values if k is odd, within the range $[-k,+k]$ in steps of 2. Its distribution is known: it is a centered (or symmetric) binomial distribution, where the probability of $S_k=2\ell-k,\quad\forall\ell=0,1,\ldots,k$ is given by

$$\mathbb{P}(S_k=2\ell-k)=\binom{k}{\ell}\left(\frac{1}{2}\right)^{\ell}\left(\frac{1}{2}\right)^{k-\ell}=\binom{k}{\ell}\left(\frac{1}{2}\right)^{k}.$$

Thus,

$$\mathbb{E}\left[|S_k|\right] = \sum_{\ell=0}^{k} |S_k| \cdot \mathbb{P}(S_k = 2\ell - k) = \frac{1}{2^k} \sum_{\ell=0}^{k} |2\ell - k| \cdot \binom{k}{\ell}$$

or, using symmetry:

$$\mathbb{E}\left[|S_k|\right] = \frac{2}{2^k} \sum_{\ell=\lceil k/2 \rceil}^{k} (2\ell - k) \cdot \binom{k}{\ell},$$

consequently,

$$\mathbb{E}\left[|u_i| \mid \mathbf{x} \neq 0\right] = \frac{1}{\sqrt{m}} \mathbb{E}\left[|S_k|\right] = \frac{2}{2^k \sqrt{m}} \sum_{\ell=\lceil k/2 \rceil}^{k} (2\ell - k) \binom{k}{\ell}. \tag{5.44}$$

(ii) Let's prove that $\mathbb{E}[|u_i| \mid \mathbf{x} \neq 0] = \frac{k}{2^{k-1}\sqrt{m}} \binom{k-1}{\lfloor (k-1)/2 \rfloor}$
We can resolve the telescoping sum in (5.44). Indeed, we can demonstrate that

$$\mathbb{E}\left[|S_k|\right] = \frac{1}{2^k} \sum_{\ell=0}^{k} |2\ell - k| \cdot \binom{k}{\ell} = \frac{k}{2^{k-1}} \binom{k-1}{\lfloor \frac{k-1}{2} \rfloor}.$$

Let us choose the case $k = 2n$ (k even), although we could also demonstrate this equivalence for the case $k = 2n + 1$ (k odd). However, for proof purposes, a single case suffices. Then for $k = 2n \Rightarrow \lceil k/2 \rceil = n$:

$$\mathbb{E}\left[|S_{2n}|\right] = \frac{2}{2^{2n}} \sum_{\ell=n}^{2n} (2\ell - 2n) \cdot \binom{2n}{\ell} = \frac{2}{2^{2n}} \sum_{\ell=n}^{2n} 2(\ell - n) \cdot \binom{2n}{\ell}$$

Setting $j = \ell - n$:

$$\mathbb{E}\left[|S_{2n}|\right] = \frac{2}{2^{2n}} \sum_{\ell=n}^{2n} 2j \cdot \binom{2n}{n+j} = \frac{4}{2^{2n}} \sum_{j+n=n}^{j+n=2n} j \cdot \binom{2n}{n+j} = \frac{4}{2^{2n}} \sum_{j=0}^{n} j \cdot \binom{2n}{n+j}$$

There exists an identity that can be demonstrated as follows:

$$\sum_{j=0}^{n} j \cdot \binom{2n}{n+j} = n \binom{2n-1}{n-1} \quad \text{(!!!)}$$

To do this, we use combinatorial identities such as the "Formule du Pion" [160, sec. 7] and by intermediate use of Pascal's formula.
Let r be a non-zero natural integer and let k be an integer such that $1 \leqslant k \leqslant r$. The "Formule du pion" identity is defined as:

$$\binom{r-1}{k-1} = \frac{k}{r} \binom{r}{k}.$$

Know that Pascal's formula is defined by

$$\binom{r}{k} = \binom{r-1}{k-1} + \binom{r-1}{k} \quad \forall (k, r) \in \mathbb{N}^2, 1 \leqslant k \leqslant r - 1$$

and with a little work manipulating the formula for the pion and Pascal's formula, one can define another identity such that:

$$\binom{r-1}{k} = \frac{r-k}{r} \binom{r}{k} \quad \forall (k, r) \in \mathbb{N}^2, 1 \leqslant k \leqslant r.$$

We take $r = 2n$, then for all k:

$$\binom{2n-1}{k-1} = \frac{k}{2n} \binom{2n}{k} \Longrightarrow k \binom{2n}{k} = 2n \binom{2n-1}{k-1} \quad \text{(A)}$$

And

$$\binom{2n-1}{k} = \frac{2n-k}{2n}\binom{2n}{k} \Longrightarrow (2n-k)\binom{2n}{k} = 2n\binom{2n-1}{k} \quad \text{(B)}$$

Subtracting (A) and (B):

$$k\binom{2n}{k} - (2n-k)\binom{2n}{k} = 2n\binom{2n-1}{k-1} - 2n\binom{2n-1}{k}$$

$$(k-2n+k)\binom{2n}{k} = 2n\left(\binom{2n-1}{k-1} - \binom{2n-1}{k}\right)$$

$$2(k-n)\binom{2n}{k} = 2n\left(\binom{2n-1}{k-1} - \binom{2n-1}{k}\right)$$

$$(k-n)\binom{2n}{k} = n\left(\binom{2n-1}{k-1} - \binom{2n-1}{k}\right)$$

Summing over $k = n, \ldots, 2n$, the telescoping sum:

$$\sum_{k=n}^{2n}(k-n)\binom{2n}{k} = \sum_{k=n}^{2n} n\left(\binom{2n-1}{k-1} - \binom{2n-1}{k}\right)$$

Letting $a_k = \binom{2n-1}{k}$, then each term equals $a_{k-1} - a_k$, so $\sum_{k=n}^{2n}(a_{k-1} - a_k) = a_{n-1} - a_{2n}$.

$$\Rightarrow \sum_{k=n}^{2n}\binom{2n-1}{k-1} - \binom{2n-1}{k} = \binom{2n-1}{n-1} - \binom{2n-1}{2n}.$$

Since $\binom{2n-1}{2n} = 0$, then multiplying again by n:

$$\sum_{k=n}^{2n}(k-n)\binom{2n}{k} = n\binom{2n-1}{n-1}$$

.

Recalling $j = k - n \Rightarrow k = n + j$:

$$\sum_{j=0}^{n} j\binom{2n}{n+j} = n\binom{2n-1}{n-1} \text{ (!!!) is proved.}$$

.

Thus:

$$\begin{aligned}\mathbb{E}\left[|S_{2n}|\right] &= \frac{4}{2^{2n}}\sum_{j=0}^{n} j \cdot \binom{2n}{n+j} = \frac{4}{2^{2n}} \cdot n\binom{2n}{n-1}, \\ &= \frac{2 \cdot 2n}{2^{2n}} \cdot \binom{2n}{n-1} = \frac{\cdot 2n}{2^{2n-1}} \cdot \binom{2n}{n-1}.\end{aligned} \tag{5.45}$$

Returning to $k = 2n$ and $\lfloor (2n-1)/2 \rfloor = n - 1$, then

$$\mathbb{E}\left[|S_k|\right] = \frac{\cdot k}{2^{k-1}} \cdot \binom{k}{\lfloor (k-1)/2 \rfloor}. \tag{5.46}$$

Therefore,

$$\mathbb{E}\left[|u_i| \mid \mathbf{x} \neq 0\right] = \frac{1}{\sqrt{m}}\mathbb{E}\left[|S_k|\right] = \frac{k}{2^{k-1}\sqrt{m}} \cdot \binom{k}{\lfloor (k-1)/2 \rfloor}. \tag{5.47}$$

■

For a more compact notation of (5.43), we define the *sufficient statistic* $T_i \coloneqq \frac{2\,\Re\{\mathrm{h}_i^* \mathrm{y}_i\}}{\sigma^2}$ and the *sparsity-dependent constant* $C_{k,m} \coloneqq \frac{k}{2^{k-1}\sqrt{m}}\binom{k-1}{\lfloor (k-1)/2 \rfloor}$.

The LLR can then be equivalently expressed as

$$\Lambda_i = C_{k,m}\, T_i, \quad i = 1, 2, \ldots, m, \quad \text{with } C_{k,m} > 0. \tag{5.48}$$

Although the weights could be defined directly from the LLR magnitudes via $\beta_i = |\Lambda_i| = C_{k,m} \cdot |T_i|$, such a straightforward formulation provides no guarantees on the range of β_i. It inherits the potentially unbounded scale of the LLRs and may suffer from numerical overflow ; specifically, $|T_i|$ can grow arbitrarily large under high signal-to-noise ratio (SNR) conditions. Hence, to obtain well-behaved weights, we employ a weighting function $f(|\mathbf{\Lambda}|)$ that maps the LLR magnitudes to weights bounded between 0 and 1, where $f(\cdot)$ denotes a suitable bounded and monotonic transformation. More importantly, the magnitude of an LLR directly reflects its reliability: larger magnitudes indicate higher decision confidence. This intrinsic reliability measure naturally justifies using LLR magnitudes to form attention weights.

Thereafter, we define the attention weights $\boldsymbol{\beta}$ from the magnitudes of the LLRs by applying a max-normalization function $f : \mathbb{R}_+ \longrightarrow [0, 1], \quad |\Lambda_i| \longmapsto f(|\Lambda_i|)$.

Formally, the weights β_i are positive and max-normalized based on the LLR magnitudes, expressed as

$$\beta_i \coloneqq f\big(|\Lambda_i|\big) = \frac{|\Lambda_i|}{\max\limits_{1 \leqslant \ell \leqslant m} |\Lambda_\ell| + \varepsilon} = \frac{C_{k,m}|T_i|}{\max\limits_{1 \leqslant \ell \leqslant m} C_{k,m}|T_\ell| + \varepsilon}, \tag{5.49}$$

where $\varepsilon > 0$ is a small regularization constant. If strictly positive weights are required, we assume $|\Lambda_i| > 0$ for all i ; hence $\beta_i \in (0, 1]$.

Subsequently, the diagonal weighting matrix is constructed as

$$\mathbf{W} = \operatorname{diag}(\beta_1, \ldots, \beta_m) = \begin{pmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_m \end{pmatrix} = \operatorname{diag}(\boldsymbol{\beta}). \tag{5.50}$$

which reweights the equalized received vector $\tilde{\mathbf{y}}$ prior to correlation with the transpose of the projection matrix.

Proposition 5.2. (LLR-Based Attention Weighted Projection.)

The decoding stage follows the procedure described in the previous subsection. In particular, the decoder evaluates the weighted correlation vector defined as

$$\mathbf{c}^{(\mathsf{w})} = \mathbf{A}^\top (\mathbf{W}, \tilde{\mathbf{y}}) \text{ where } \mathrm{c}_j^{(\mathsf{w})} = \mathbf{a}_j^\top (\boldsymbol{\beta} \odot \tilde{\mathbf{y}}) = \langle \mathbf{a}_j, \boldsymbol{\beta} \odot \tilde{\mathbf{y}} \rangle. \tag{5.51}$$

The support of $\mathbf{x}$ is then estimated by selecting the indices corresponding to the k largest-magnitude entries of $\mathbf{c}^{(\mathsf{w})}$. Figure 5.7 provides a conceptual schematic of the SVC receiver employing LLR-based attention-weighted projection decoding, while Algorithm 5 summarizes the associated decoding procedure.

Corollary 5.2. *(Weighted Coherence and Minimal Diagonal Element.)*

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{W} = \operatorname{diag}(\boldsymbol{\beta}) \geqslant 0$ and $\mathbf{G} = \operatorname{diag}\left(\frac{|\mathrm{h}_1|^2}{|\mathrm{h}_1|^2+\varepsilon}, \ldots, \frac{|\mathrm{h}_m|^2}{|\mathrm{h}_m|^2+\varepsilon}\right) \geqslant 0$.

The weighted coherence $\mu_{\mathbf{W}}$ satisfies:

$$\mu_{\mathbf{W}} = \max_{i \neq j} \left|\mathbf{a}_i^\top \mathbf{W}\mathbf{G}\mathbf{a}_j\right|, \tag{5.52}$$

and the minimal diagonal by

$$\alpha_{\min} = \min_j \mathbf{a}_j^\top \mathbf{W}\mathbf{G}\mathbf{a}_j > 0. \tag{5.53}$$

$\alpha_{\min} > 0$ represents a measure of the smallest effective diagonal value via $\mathbf{a}_j$, the vectors in the matrix combination $\mathbf{WG}$. This ensures that all directions have a positive contribution and limits negative interference.

In other respects, if the weighted coherence $\mu_{\mathbf{W}}$ is low and the minimum diagonal element of $\mathbf{A}^\top\mathbf{WGA}$ is large compared to the noise term, then the top-k selection recovers the support exactly with high probability. We then argue that the LLR-based weighting reduces $\mu_{\mathbf{W}}$ because it downweights measurements with low channel gain, which are more prone to noise.

Proof. (Weighted coherence and minimal diagonal element)

Expanding $\mathbf{c}^{\mathsf{weighted}}$ expression as

$$\begin{aligned}\mathbf{c}^{\text{weighted}} &= \mathbf{A}^\top \mathbf{W}\tilde{\mathbf{y}} = \mathbf{A}^\top \mathbf{W}\left(\Re\left(\frac{\mathbf{h}^* \odot \mathbf{y}}{\|\mathbf{h}\|^2 + \varepsilon}\right)\right), \\ &= \mathbf{A}^\top \mathbf{W}\mathbf{G}\mathbf{A}\mathbf{x} + \mathbf{A}^\top \mathbf{W}\tilde{\mathbf{z}}.\end{aligned} \tag{5.54}$$

Given that the deterministic component of $\mathbf{c}^{\text{weighted}}$, which corresponds to the matrix $\mathbf{A}^\top\mathbf{W}\mathbf{G}\mathbf{A}$ has off-diagonal elements that may cause interference, the the weighted coherence is then governed by this term as which is simply $\mu_{\mathbf{W}} = \max_{i\neq j} \left|\mathbf{a}_i^\top \mathbf{W}\mathbf{G}\mathbf{a}_j\right|$

■

Algorithm 5 summarizes the LLR-based attention-weighted decoding steps.

Algorithm 5: LLR-Based Attention-Weighted Decoding.

Require: Projection matrix $\mathbf{A}$, received vector $\mathbf{y}$, channel diagonal matrix $\mathrm{diag}(\mathbf{h})$, sparsity level k, Noise variance σ^2 .

1: Equalized received vector: $\tilde{\mathrm{y}}_i = \Re\left(\frac{\mathrm{h}_i^* \mathrm{y}_i}{|\mathrm{h}_i|^2+\varepsilon}\right)$.
2: Compute LLRs: $\Lambda_i = \frac{2\Re\{\mathrm{h}_i^* \mathrm{y}_i\}}{\sigma^2} \cdot \mathbb{E}[|\mathbf{a}_i^\top \mathbf{x}| \mid \mathbf{x} \neq 0]$.
3: Compute attention weights: $\beta_i = \frac{|\Lambda_i|}{\max_\ell |\Lambda_\ell| + \varepsilon}$.
4: Form diagonal weighting matrix $\mathbf{W} = \mathrm{diag}(\boldsymbol{\beta})$.
5: Compute weighted correlation: $\mathbf{c}^{(\mathrm{w})} = \mathbf{A}^\top(\mathbf{W}\tilde{\mathbf{y}})$.
6: Estimate support set: $\widehat{\mathcal{S}} = \mathrm{top}_k\left(|\mathbf{c}^{(\mathrm{w})}|\right)$.
7: Reconstruct sparse vector $\hat{\mathbf{x}}$ with:
$\hat{\mathrm{x}}_j = \mathrm{sign}(\mathrm{c}_j^{(\mathrm{w})})$ for $j \in \widehat{\mathcal{S}}$, and $\hat{\mathrm{x}}_j = 0$ elsewhere.
8: **return** $\hat{\mathbf{x}}$.

Figure 5.7 presents a conceptual illustration of the SVC receiver using LLR-based attention-weighted projection decoding.

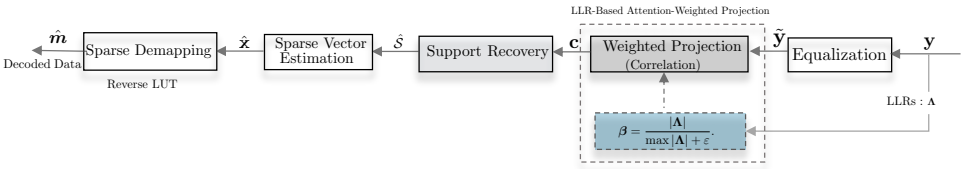


Figure 5.7: Conceptual illustration of the SVC receiver employing LLR-based attention-weighted projection decoding.

3.3.3 Exhaustive Maximum Likelihood Decoding as Benchmark

The sparse vector $\mathbf{x} \in \mathbb{R}^n$ can be recovered via exhaustive Maximum Likelihood (ML) decoding. Combined with the LLR-based attention weighting principle, the proposed decoder exhibits clear advantages over the baseline approach. Although the baseline ML algorithm is characterized by rapidly increasing computational complexity, it remains a reference solution when optimal performance is required. Hence, it serves as a solid benchmark for comparison purposes.

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathcal{X}}{\mathrm{argmin}} \|\tilde{\mathbf{y}} - \mathbf{A}\mathbf{x}\|^2, \quad \forall \mathbf{y} \in \mathbb{R}^m, \quad \forall \mathbf{A} \in \mathbb{R}^{m\times n}. \tag{5.55}$$

Expanding the quadratic term yields

$$\|\tilde{\mathbf{y}} - \mathbf{A}\mathbf{x}\|^2 = \|\tilde{\mathbf{y}}\|^2 + \|\mathbf{A}\mathbf{x}\|^2 - 2\tilde{\mathbf{y}}^\top \mathbf{A}\mathbf{x}.$$

The term $\|\tilde{\mathbf{y}}\|^2$ is constant and independent of $\mathbf{x}$, and can therefore be omitted. Consequently, the maximization with respect to $\mathbf{x}$ depends only on the remaining two terms:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathcal{X}}{\mathrm{argmax}}\ 2\tilde{\mathbf{y}}^\top \mathbf{A}\mathbf{x} - \|\mathbf{A}\mathbf{x}\|^2. \tag{5.56}$$

At the message-bit recovery level via ML decoding, it is not necessary to explicitly estimate the sparse vector $\hat{\mathbf{x}}$ beforehand. Instead, the binary message $\hat{\boldsymbol{m}}$ can be directly estimated through exhaustive comparison over the $|\mathcal{M}| = 2^{\lfloor \log_2 \binom{n}{k} \rfloor + k}$ candidate messages:

$$\hat{\boldsymbol{m}} = \underset{\boldsymbol{m} \in \mathcal{M}}{\operatorname{argmax}} \, 2\tilde{\mathbf{y}}^\top \mathbf{A}\mathbf{x}(\mathbf{m}) - \|\mathbf{A}\mathbf{x}(\boldsymbol{m})\|^2 . \tag{5.57}$$

3.4 SVC for 3GPP PUCCH Formats 2, 3, 4

In the 3GPP standard, UCI packets in the 3-11 bit range are conveyed on PUCCH using Reed Muller encoding, followed by rate matching, scrambling, complex-valued QPSK modulation, in general, even if complex-valued $\pi/2$ BPSK modulation is allowed for PUCCH format 3 / 4. The resulting complexe-modulated symbols are then mapped to subcarriers across multiple resource blocks using one or multiple OFDM symbols. For channel estimation, a pseudo-random QPSK sequence is used as a DMRS (Demodulation Reference Signal), enabling the base station to resolve channel ambiguities in time, frequency, or space before performing coherent detection. Furthermore, the principle of Reed-Muller-coded transmission for the 3GPP PUCCH is detailed in [132].

Nevertheless, it appears increasingly clear that 3GPP Reed–Muller codes, as currently designed to be decoded using well-known maximum likelihood decoding, are not optimal in certain respects, particularly from a computational complexity standpoint. Decoding via exhaustive ML search is effectively prohibitive, especially in the context of short-packet transmissions targeting URLLC use cases. Therefore, in what follows, we seek to address the following question: how can sparse-aware vector-coded transmission emerge as a viable alternative, potentially rivaling or even surpassing standard 3GPP RM-coded transmission in terms of both performance and decoding complexity?

3.4.1 3GPP PUCCH Format 2

PUCCH Format 2 is a short, OFDM-based format capable of carrying more than two bits of uplink control information (UCI) over just one or two OFDM symbols. Its transmission structure is relatively simple. Within a slot, the DMRS (Demodulation Reference Signal) sequences are frequency-division multiplexed with the data on either the last one or two OFDM symbols. This sequence is mapped to every third subcarrier within each OFDM symbol.

3.4.2 3GPP PUCCH Format 3 / 4

PUCCH format 3 is designed for coverage-limited scenarios, unlike format 2, which is better suited for small cells. To improve coverage, it achieves higher received energy by spreading the information across more OFDM symbols. It also employs DFT-precoding (SC-FDMA) to maintain a low peak-to-average power ratio. With the ability to transmit more than two bits using between 4 and 14 symbols, this format offers one of the largest payload capacities among the PUCCH options.

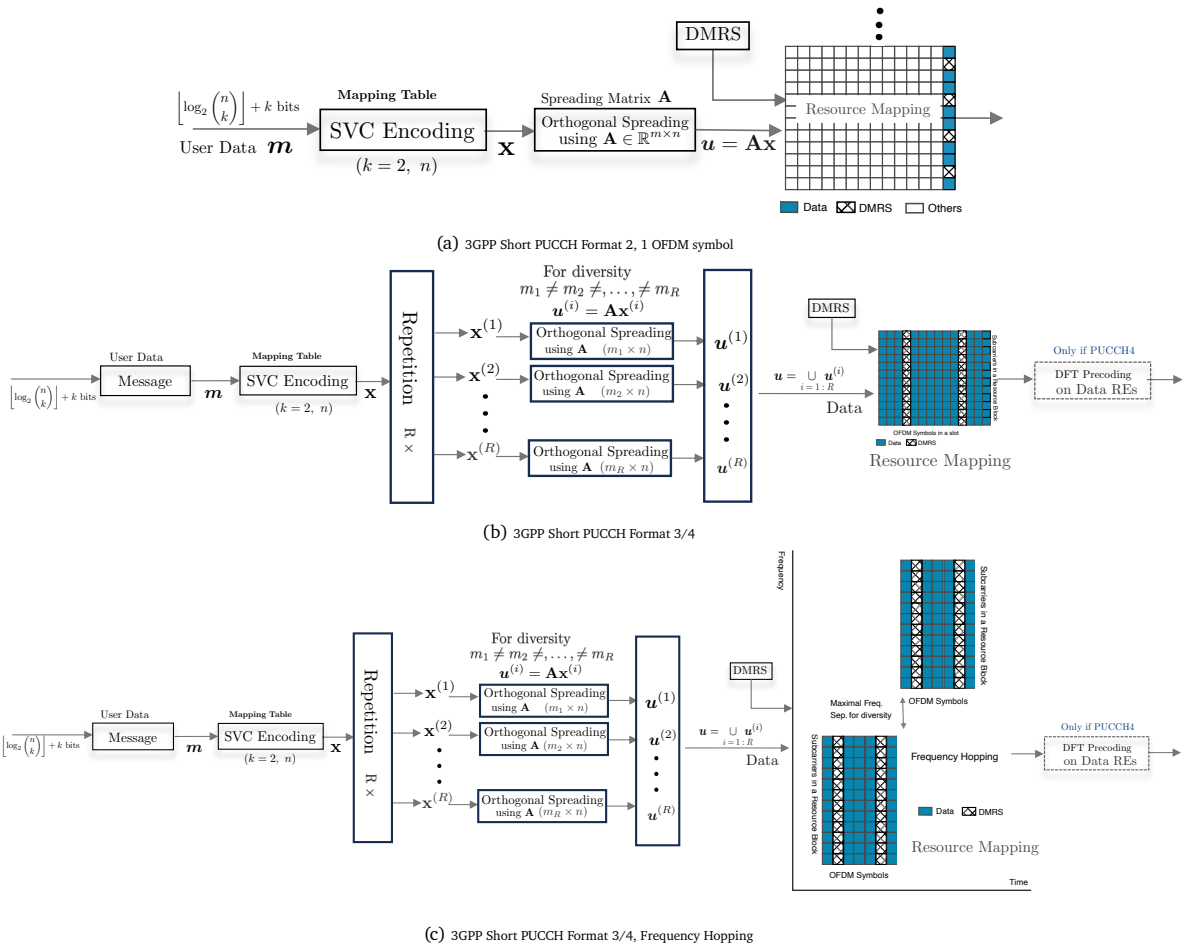
Additionally, PUCCH format 4's structure is similar to format 3, but it adds the ability to code-multiplex multiple devices on the same resource block. Within each OFDM symbol, $12/N_{\mathsf{SF}}$ unique modulation symbols are transmitted. Before DFT-precoding, each symbol is block-spread using an orthogonal sequence of length N_{SF}. Since the supported spreading factors are 2 and 4, this allows either two or four devices to share the same set of resource blocks [18, p.615]. The DMRS sequences are mapped to all subcarriers in the fifth and tenth OFDM symbols of a slot.

Although frequency hopping is optional for PUCCH formats 3 and 4, it can be configured to take advantage of frequency diversity. Furthermore, it is possible to consider a spreading matrix $\mathbf{A}$ of different dimensions for each repetition in order to introduce some diversity. In other words, for each repetition r, a spreading matrix $\mathbf{A}$ with m_r rows is used. The key point is that the total number of elements combining the sub-measurement vectors $\mathbf{u}_i$ corresponds to the number of REs allocated to the data. With this diversity, we will have a better chance of recovering the support of the sparse vector and therefore of finding the correct message with more probability. Figure 5.8 presents a conceptual illustration of sparse aware vector coded transmission adapted to the 3GPP PUCCH structure, using PUCCH Format 2/3/4 as representative instances, specifically in the resource mapping fashion.

3.5 Numerical Results

3.5.1 Proposed SVC vs Conventional approcaches

For illustrative purposes, we consider an SVC transmission with a sparse vector length $n = 64$, sparsity level $k = 2$, and a spreading matrix $\mathbf{A} \in \mathbb{R}^{32 \times n}$. This configuration operates in the short blocklength regime and corresponds to an information message $\boldsymbol{m}$ carrying a payload of $B = 12$ bits, which lies at the lower end of the short-packet communication range, typically on the order of a few tens of bits. The performance comparison includes: (– –⊗– –)



(a) 3GPP Short PUCCH Format 2, 1 OFDM symbol

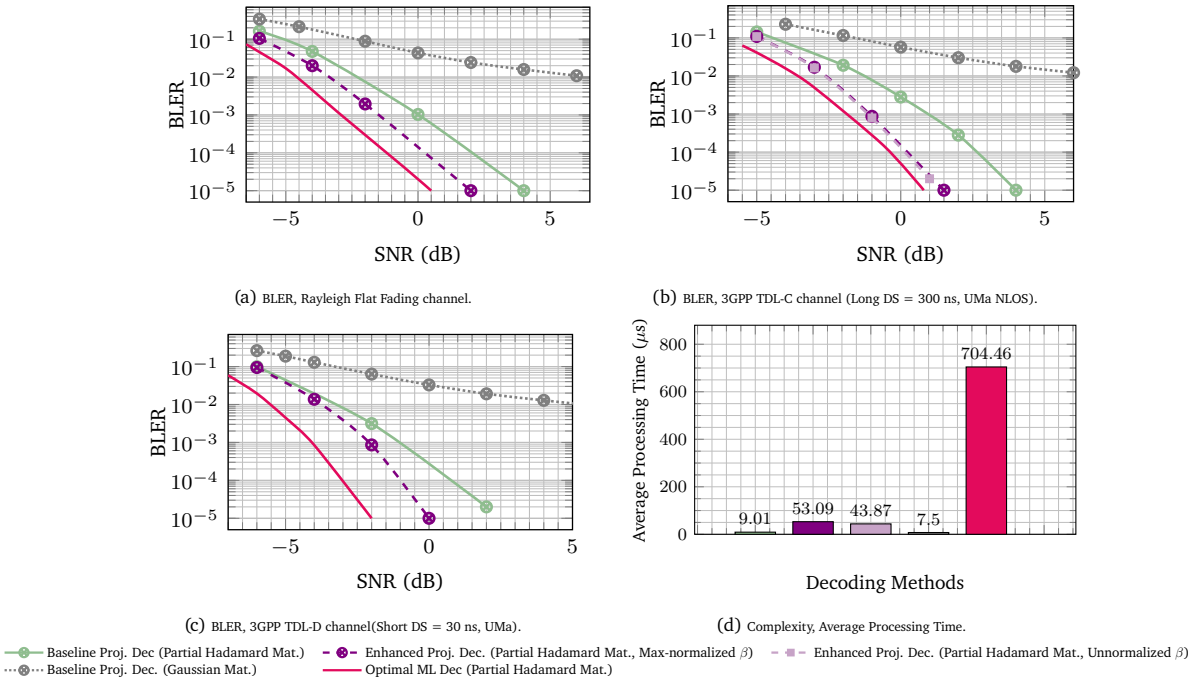
(b) 3GPP Short PUCCH Format 3/4

(c) 3GPP Short PUCCH Format 3/4, Frequency Hopping

Figure 5.8: Conceptual illustration of sparse vector coded transmission for 3GPP PUCCH (e.g.,Formats 2, 3 & 4).

the proposed *enhanced projection decoding* based on an LLR-based attention-weighted approach, (—⊗—, · · ⊗ · ·) the *baseline projection decoding* employing either a Hadamard projection matrix or a random Gaussian matrix, and the optimal ML decoder (—).

In addition, the performance evaluation in this subsection assumes perfect channel state information ; however, the analysis can be straightforwardly extended to scenarios wherein the channel state information is unknown. In such cases, a pilot-assisted transmission scheme can be employed, which will also be illustrated in the subsequent subsection for 3GPP PUCCH using DMRS-assisted transmission. Here, in the scenario of interest, SIMO diversity is exploited using multiple receive antennas in conjunction with the maximal ratio combining (MRC) algorithm, which performs optimal weighted combining based on the channel gains to enhance transmission reliability by mitigating fading and improving the SNR. Both Rayleigh flat fading channels (Figure 5.9a) and frequency-selective channels modeled using the 3GPP tapped delay line (TDL) models, specifically TDL-C (long delay spread of 300 ns) and TDL-D (short delay spread of 30 ns) for urban macro (UMa) scenarios [135] are considered, as illustrated in Figures 5.9b and 5.9c, respectively. As previously stated, TDL-C and TDL-D capture multipath propagation with distinct delay profiles and are tailored for non-MIMO channel assessments, thus enabling a realistic evaluation of the proposed decoding scheme under practical propagation conditions. In this multiple-receive-antenna scenario, the attention weights are computed by averaging the β values across the N_R receive antenna branches. It is noteworthy that the LLR-based attention-weighted decoder exhibits superior performance under challenging channel conditions characterized by severe fading, noise, or interference. For instance, in the TDL-C NLOS scenario, signal dispersion over multiple weak multipath components complicates detection at low SNR, as the receiver must reconstruct information from fragmented and noisy signals. At a very low BLER threshold of 0.001%, the proposed enhanced decoder—based on LLR-based attention-weighted Hadamard projection demonstrates an approximate 2 dB gain over the baseline Hadamard projection decoder. These gains and performance differences are consistently observable across all three channel types considered, as shown in Figure 5.9. Importantly, performance critically depends on the properties



(a) BLER, Rayleigh Flat Fading channel.

(b) BLER, 3GPP TDL-C channel (Long DS = 300 ns, UMa NLOS).

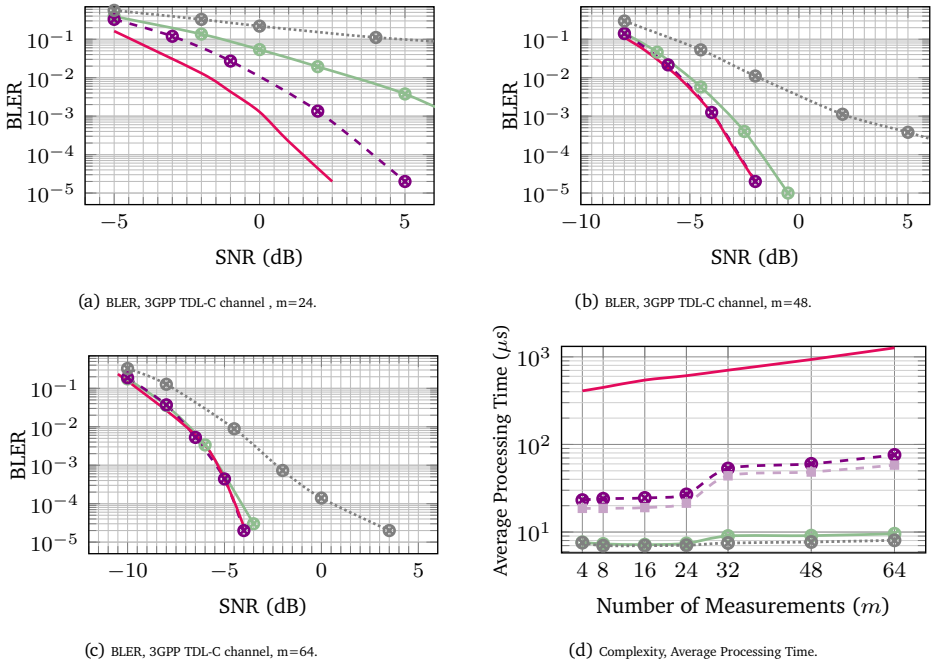
(c) BLER, 3GPP TDL-D channel(Short DS = 30 ns, UMa).

(d) Complexity, Average Processing Time.

Figure 5.9: Block error rate (BLER) and average processing complexity analysis. **SVC parameters:** sparse vector length $n = 64$, sparsity $k = 2$, spreading matrix $\mathbf{A} \in \mathbb{R}^{32\times 64}$, message $\boldsymbol{m}$ of 12 bits ($\boldsymbol{m}_1 = 10$ bits + $\boldsymbol{m}_2 = 2$ bits), coding rate $12/64$. **Decoders:** baseline partial Hadamard projection, baseline random Gaussian projection, and LLR-based attention-weighted Hadamard projection. **Channel configurations:** perfect CSI, 1×4 SIMO with (a) Rayleigh flat fading, (b) 3GPP TDL-C NLOS (long delay spread = 300 ns), and (c) 3GPP TDL-D LOS+NLOS (short delay spread = 30 ns); sampling rate $f_s = 30.72$ MHz.

of the projection matrix, particularly its orthogonality and robustness to noise, fading, and interference. Notably, configurations employing random Gaussian projection decoding consistently exhibit the poorest performance, with an error floor around 1% BLER. In contrast, configurations utilizing a partial Hadamard matrix achieve significantly lower BLER, with the LLR-based attention-weighted projection decoding providing an additional performance improvement, nearing the optimal performance bound achieved by ML decoding. Moreover, computational complexity, assessed via average processing time on a semi-logarithmic scale, is reported in Figure 5.9d. The proposed decoder is evaluated both with weights formed from max-normalized and unnormalized LLR magnitudes, and is compared against the baseline and optimal ML decoders. The enhanced decoder cycle comprises LLR computation, weight formulation, and projection calculation, simulated over 10^4 Monte Carlo trials at a given SNR. Relative to the baseline, the proposed decoder incurs approximately a $6\times$ increase in processing time ; however, it remains nearly $13\times$ less complex than full ML decoding. Notably, both the unnormalized and max-normalized LLR weighting configurations yield identical BLER performance. Nonetheless, the unnormalized variant is computationally more efficient, as the max-normalization step introduces an additional overhead of approximately 10 μs per decoding instance. Furthermore, Figure 5.10 illustrates the BLER performances and average processing time complexity as a function of the measurement dimension m. It is worth noting that the proposed receiver achieves higher gains over the baseline counterpart when the measurement dimension satisfies $m \lesssim n/2$. For larger dimensions, particularly as m approaches n (i.e., when a full Hadamard matrix is employed) and under perfect CSI conditions, the performance improvement relative to baseline projection becomes negligible. Specifically, when $m = n$ under perfect CSI, the performance of the proposed projection decoder and its baseline counterpart is identical, as the full orthogonality of the Hadamard projection matrix is already fully exploited in both cases. From a complexity perspective, the computational cost increases with the number of measurements for all benchmark decoders. Hence, for low-latency applications, a smaller measurement dimension m should be adopted, while the attention-weighting mechanism can compensate for the performance gap relative to the optimal ML performance bound.

Figure 5.11 illustrates the average attention weight β as a function of SNR and its probability density functions for low (-10 dB), medium (0 dB), and high ($+10$ dB) SNR regimes. The attention weight β, computed via max-normalization of LLR magnitudes, increases monotonically with SNR, demonstrating its effectiveness as a channel-aware



(a) BLER, 3GPP TDL-C channel , m=24.

(b) BLER, 3GPP TDL-C channel, m=48.

(c) BLER, 3GPP TDL-C channel, m=64.

(d) Complexity, Average Processing Time.

Figure 5.10: Block Error Rate (BLER) and average processing complexity for SVC with $n = 64$, $k = 2$, varying measurement dimension m, and 12-bit messages. Perfect CSI, 1×4 SIMO TDL-C NLOS channel (long delay spread = 300 ns, $f_s = 30.72$ MHz).

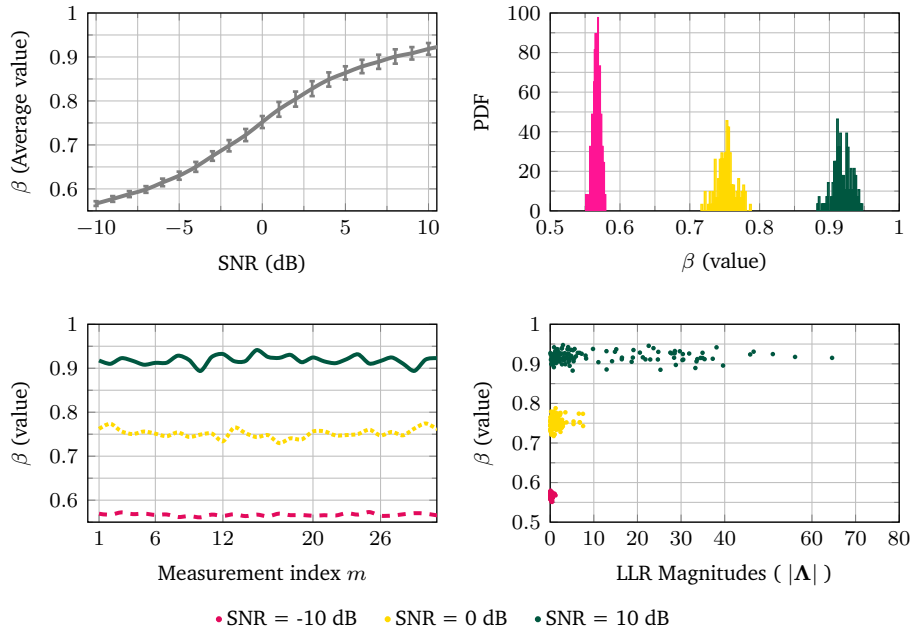


Figure 5.11: Analysis of attention weight behavior in sparse recovery. **(a)** Evolution of average attention weight β versus SNR, exhibiting monotonic increase with channel quality and (b) its Probability density functions (c) β as a function of the measurement index, and (d) β versus LLR magnitudes over low (-10 dB), medium (0 dB), and high ($+10$ dB) SNR regimes.

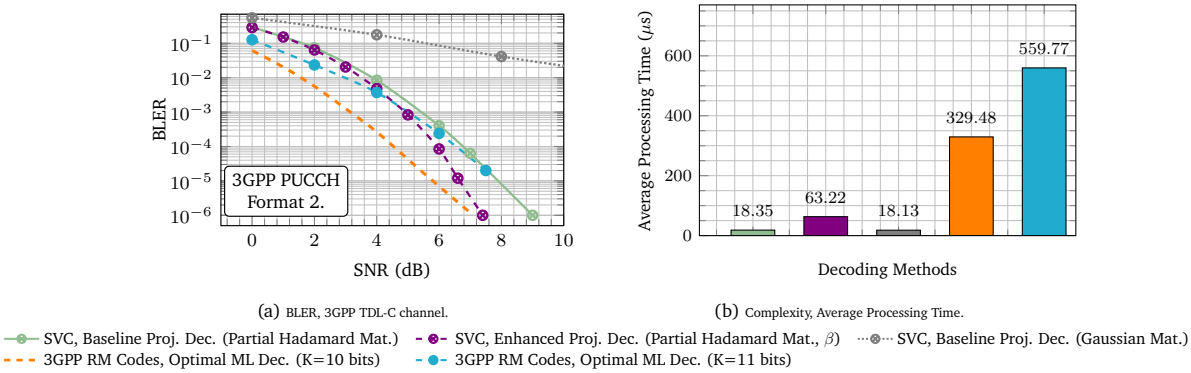
reliability metric for sparse signal recovery. The figure shows that the distribution of attention weights β strongly depends on the SNR. At low SNR (-10 dB), the weights are concentrated around small values (approximately 0.55), whereas at high SNR ($+10$ dB), they are centered around larger values (approximately 0.92). At low SNR (e.g., -10 dB), the β_i values are generally lower and more uniform, since few measurements exhibit high LLR values. Conversely, at high SNR (e.g., 5 dB), the variation among the β_i values becomes more pronounced and the average increases, reflecting a clearer distinction between reliable and unreliable measurements. Furthermore, although the empirical

PDF of β appears more peaked at low SNR, this behavior reflects a lack of discriminative power, as most weights collapse to similar values due to noise-dominated LLRs. In contrast, at high SNR, the broader distribution indicates increased variability and selectivity of the attention mechanism, enabling more effective support estimation and sparse-aware signal reconstruction.

3.5.2 Proposed SVC vs 3GPP RM Coding over PUCCH 2

As described above, for illustrative purposes, we consider a sparse vector-coded transmission on a 3GPP PUCCH Format 2 with $n = 32$, $k = 2$, and measurement dimension $m = 32$. The information message $\boldsymbol{m}$ has a payload of $K = 10$ bits, which falls within the 3–11 bit short-packet range specified by the 3GPP standard for RM-coded transmissions. A payload of $K = 11$ bits is also considered for additional performance analysis.

Accordingly, we consider a scenario with unknown channel conditions, and thus a DMRS-assisted transmission configuration is deployed. A single OFDM symbol spans four physical resource blocks (PRBs), in which 32 resource elements (i.e., m elements of the measurement vector after orthogonal spreading) are allocated to data REs, while the remaining 16 REs are reserved for demodulation reference signals (DMRSs). Channel estimation is performed using the least squares method, followed by linear interpolation to track the channel at the data positions. Consistent with previous experiments, we adopt the 3GPP TDL-C NLOS channel model with a long delay spread of 300 ns. The performance evaluation considers three schemes: the proposed *SVC with enhanced projection decoding*, the *SVC with baseline projection decoding* employing either a Hadamard projection matrix or a random Gaussian matrix, and *3GPP RM codes with optimal ML decoding* (– – –) serving as a benchmark. The optimal ML decoder is implemented via exhaustive search. For a fair comparison, the same number of message bits is maintained across both coding schemes, i.e., K bits (3GPP RM) $\equiv$ $B = \lfloor \log_2 \binom{n}{k} \rfloor + k$ bits (SVC). As reported in the previous subsection, the



(a) BLER, 3GPP TDL-C channel. (b) Complexity, Average Processing Time.

Figure 5.12: Block Error Rate comparison for 3GPP PUCCH Format 2 transmission. **PUCCH Configuration:** 1 OFDM symbol, 4 PRBs, 32 data REs, 16 DMRS REs. **SVC Parameters:** sparse vector length $n = 32$, sparsity $k = 2$, message $\boldsymbol{m}$ of $B = 10$ bits, coding rate $10/32$. **3GPP Reed–Muller:** $\mathscr{C}(32, 10)$ code, message of $K = 10$ bits, coding rate $10/32$, QPSK modulation. **Decoders:** baseline Hadamard projection vs. LLR-based attention-weighted Hadamard projection vs. ML decoding for RM codes. **Channel configurations:** unknown CSI, LS estimation with linear interpolation, 1×4 SIMO TDL-C NLOS channel (long delay spread = 300 ns, $f_s = 30.72$ MHz).

conventional approach using *SVC with a Gaussian random spreading matrix and baseline projection decoding* remains less performant than all other configurations considered in Figure 5.12. In contrast, SVC with a Hadamard spreading matrix combined with LLR-based attention-weighted projection decoding achieves performance comparable to 3GPP RM-coded transmission under optimal ML decoding at very low BLER targets. For example, at a BLER of 0.0001%, both schemes demonstrate equivalent performance on PUCCH Format 2. As the BLER target becomes more stringent, the proposed SVC receiver is expected to outperform the conventional ML receiver for 3GPP RM codes. Furthermore, from a computational complexity perspective, as depicted in Figure 5.12b, the proposed decoder offers approximately a $5\times$ reduction in processing time compared to 3GPP RM under ML decoding, thereby contributing to the fulfillment of URLLC requirements.

Furthermore, an additional performance assessment involving a payload of $K = 11$ bits over 3GPP RM-coded transmission shows that computational complexity increases almost exponentially with the payload size, whereas SVC maintains a more moderate growth in computational complexity relative to the payload.

As a whole, given that SVC does not require long codes or complex modulation schemes to achieve a satisfactory trade-off between performance, latency, and complexity, it appears to be a viable coding choice for short-packet

transmission in next-generation wireless standards.

3.5.3 Perspective and Future Work

Furthermore, as a perspective, the proposed SVC framework, implemented at both the transmitter and receiver, can be extended to incorporate an outer error-correcting code (ECC) to enhance reliability, thereby providing an additional layer of protection. Based on the results presented above, the SVC, particularly with the LLR-based attention-weighted projection can achieve very low BLER even under highly challenging scenarios. In such cases, the inclusion of an outer ECC may be unnecessary, as it would introduce additional complexity. Nevertheless, for more critical applications, the use of a conventional outer ECC could be advantageous. In this configuration, the SVC would operate as a higher-layer coder–modulator, combined with an ECC serving as a dedicated protection layer.

4 Summary

In this chapter, we have presented the analysis and performance enhancement of short block-length channel transmission assisted by training-based information or reference signals. 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. Hence, it was demonstrated that employing an adaptive DMRS/Data power adjustment can reduce the performance loss gap, facilitating a more sensitive and lower complexity quasi-coherent receiver. However, both of the aforementioned ML detection strategies may incur prohibitive computational complexity when processing long bit length codes within the SBL regime. To address this challenge, we proposed an innovative approach by introducing the principle of block/segment coding using *First-order RM* (FoRM) codes, which are amenable to low-cost decoding through *block-based fast Hadamard transforms* (BFHT). Therefore, through the incorporation of an adaptive DMRS/Data power adjustment technique, we were able to bridge or reduce the performance gap compared to the conventional maximum likelihood receiver and achieve high sensitivity, resulting in a favorable trade-off between performance and complexity to efficiently handle small payloads.

Furthermore, the second part of this chapter discusses another paradigm for short packet transmission. We have presented an enhanced SVC scheme tailored for short-packet communication over noisy wireless channels. We have shown that conventional SVC approaches relying on random (Gaussian, Bernoulli) or structured (partial Hadamard) spreading/projection matrices with baseline projection decoding may suffer from reliability degradation under fading conditions. To overcome this limitation, we have introduced an enhanced decoding approach leveraging LLR-based attention-weighted projection, which dynamically reweights received signal measurements according to their reliability, giving more importance to the most reliable measurements. Through simulations over both Rayleigh flat fading and 3GPP TDL-C/TDL-D frequency-selective channels, we have demonstrated that the proposed LLR-based attention-weighted projection decoding outperforms baseline projection decoding with both structured and random projection matrices. Furthermore, we have extended the analysis to a short-packet (3–11 bits) transmission framework over the 3GPP PUCCH. The results indicate that the proposed SVC scheme, when combined with a partial Hadamard spreading matrix and LLR-based attention-weighted projection decoding, can competitively rival 3GPP RM codes under optimal ML decoding at very low BLER targets, while exhibiting significant computational complexity advantages. Overall, these findings make the proposed scheme a promising short-packet coding candidate for ultra-reliable low-latency beyond 5G/6G short-block-length UL/DL channels.

Chapter 6

Closure and Outlook

> Man is in a hurry, and time is long,
> but everything has its time.
>
> — Djibril Tamsir Niane

In response to the stringent requirements for high reliability and low latency in beyond-5G systems, advanced transmission approaches and strategies have been the focus of considerable research efforts in recent years. They primarily rely on two key criteria: improving performance and reducing the complexity of algorithms and protocols supporting communication systems, particularly those involving short packet transmission, to push the fundamental boundaries in terms of coded signal design and advanced receivers. The objectives of the research developed in this thesis were to propose new strategies for designing sophisticated and advanced receivers for *short data transmissions* to meet reliability and low latency requirements. Short packet transmission involves advanced coded modulation schemes including state-of-the-art channel codes, namely Polar, LDPC, and Reed Muller codes. The defined objectives were addressed in two phases: first, receiver design for 5G polar/LDPC coded modulation systems, and second, receiver design for 5G Reed Muller coded transmission systems.

In Chapters 3 and 4, we presented novel coded modulation receiver metrics for joint estimation–detection in short blocklength transmissions, specifically targeting scenarios with unknown channel state information and low or sparse training resources. We demonstrated the potential to enhance the performance and sensitivity of advanced MIMO and non-MIMO receivers, particularly when channel state information is unknown and the density of training dimensions is low. In this respect, through block-wise joint estimation–detection, we demonstrated significant improvements in performance and sensitivity compared to conventional receivers. Our analysis, conducted using full 5G transmitter and receiver chains with both Polar and LDPC coded transmissions under QPSK modulation, shows that even when reference signals are interleaved with coded data over a small number of OFDM symbols, precluding near-perfect channel estimation. Unlike conventional symbol-by-symbol detection in BICM systems, where the observation for a given coded bit is confined to the symbol in which it is conveyed,the proposed method performs block-wise joint detection over a sliding window of adjacent symbols to fundamentally leverages their statistical dependencies; that is, the LLR for each coded bit incorporates information from multiple symbols rather than being confined to its host symbol. Performance evaluation spans SIMO and SU-MIMO configurations, emphasizing the efficacy of the estimation-detection strategy 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 receivers remarkably outperform the conventional ones and can be used to achieve detection performance that is close to that of coherent receivers with perfect CSI via DMRS power boosting for both polar and LDPC coded transmissions.

Subsequently, in Chapter 5, we presented a comprehensive analysis and performance improvement of short block channel detection (3-11 bits) within Reed Muller coded transmission. By initially investigating the structure of the maximum likelihood receiver in short data transmissions, it was demonstrated that the non-coherent energy term, which is often neglected in conventional detection metrics, leads to significant losses in terms of performance and sensitivity in typical operating regimes of 5G/6G systems. Comparison of this conventional or quasi-coherent receiver

with the fully non-coherent receiver in multi-antenna configurations revealed that these substantial losses become more pronounced in low spectral efficiency areas. Furthermore, it was proved that by using adaptive adjustment of DMRS/Data power, the performance loss gap may be reduced, resulting in a more sensitive quasi-coherent receiver with minimal complexity. However, it appears that the aforementioned ML detection strategies may result in significant computational complexity when processing long bit length codes within this short block regime, since both rely on an ML receiver. Therefore, we subsequently proposed an approach to address 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. Block-based FHT has proven to be cost-effective in terms of decoding time, as it moves from quadratic complexity to quasi-linear complexity with a manageable loss of performance. By incorporating an adaptive DMRS power adjustment technique, we can bridge/reduce the underlying performance gap and achieve high sensitivity, resulting in a good trade-off between performance and complexity for efficiently processing small payloads. Alongside this framework, in the search for alternative codes that could compete with or even outperform 3GPP RM codes, we are relying on the sparse vector coding (SVC) approach. The SVC scheme is simple, flexible, and of low complexity, making it suitable for short packet transmissions. The baseline decoding process is projection based, following a simplified Orthogonal Matching Pursuit. Hence, we have proposed an enhanced projection decoding approach leveraging LLR-based attention-weighted projection, which dynamically reweights received signal measurements according to their reliability, giving more importance to more reliable measurements. Through simulations over both Rayleigh flat fading and 3GPP TDL-C/TDL-D frequency-selective channels, we have demonstrated that the proposed LLR-based attention-weighted projection decoding outperforms baseline projection decoding with both structured and random projection matrices. Furthermore, we have extended the analysis to a short-packet (3–11 bits) transmission framework over the 3GPP PUCCH. The results indicate that the proposed SVC scheme, when combined with a partial Hadamard spreading matrix and LLR-based attention-weighted projection decoding, can competitively rival 3GPP RM codes under optimal ML decoding at very low BLER targets, while exhibiting significant computational complexity advantages, making it a promising short-packet coding candidate for ultra-reliable low-latency 6G short-block-length UL/DL channels.

This thesis opens up several perspectives that we will strive to address in future work. Although innovative BICM receiver designs have been proposed with sustained performance gains compared to conventional receivers in Chapters 3, 4, and 5, extensions and improvements are still possible and require further investigation. The proposed block-wise joint estimation–detection approach has inherent limitations. It becomes computationally prohibitive for higher-order modulation schemes such as 16-QAM, since with a detection block size of $M = 4$, this would require processing $|\boldsymbol{\chi}|^M = 65,536$ possible symbol vector combinations, which is computationally intractable. In such scenarios, reducing the detection block size (e.g., to $M = 2$) becomes necessary to maintain feasible complexity. Investigating scenarios involving high-order modulations in which the detection block size is reduced (e.g., to $M = 2$) in order to preserve manageable computational complexity remains an avenue for exploration. Moreover, the proposed JED-based receivers may be extended to incorporate iterative processing to improve LLR reliability and enhance overall transmission quality.

Accordingly, the proposed receiver metrics can be extended and refined to account for other block schemes (>100 bits) ranging from medium to long, with a greater emphasis on managing complexity in these cases. However, within the scope of this work, the studies have been limited to specific usage scenarios involving a single user. We should consider more general cases, particularly by exploiting transmission over synchronous and/or asynchronous multiple access channels. Studying these scenarios seems realistic because, unlike the transmissions described herein, transmitters may be either completely or partially desynchronized from receivers, combined with a probable lack of knowledge or uncertainty about channel state information. Indeed, short packet transmission (e.g., payload on MsgA-PUSCH in 2-step RA) could involve a multiple access channel with an unknown number of simultaneous transmitters. In this sense, the signal code and receiver design problems introduce another significant level of complexity by combining both additional channel uncertainty and multiple access, and will be a central area of investigation in future research efforts.

As a whole, these contributions advance the state of the art and push the boundaries of receiver algorithm design. The methods and solutions proposed in this manuscript demonstrate that using robust, lower-complexity advanced receivers for enhanced short data communications, towards beyond 5G URLLC use cases, appears to be tractable.

Bibliography

[1] ETSI, "5G", European Telecommunications Standards Institute, Accessed: 2024-04-14, https://www.etsi.org/technologies/5g.

[2] C. E. Shannon, “A mathematical theory of communication,” Bell Labs Technical Journal, vol. 27, pp. 379 – 423, 623 – 656, October 1948.

[3] R. Hamming, ”Error detecting and error correcting codes," Bell System technical journal, vol. 29, pp. 147-160, 1950.

[4] C. Berrou, ”Error-correction coding method with at least two systematic convolutional codings in parallel, corresponding iterative decoding method, decoding module and decoder," 1995.

[5] S. . R. Whitaker, J. Canaris et K. Cameron, ”Reed-Solomon VLSI Codec for advanced television," IEEE Trans. Circuits and Systems for video tech, vol. 1, 1991.

[6] A. Hocquenghem, ”Codes correcteurs d’erreurs," Chiffres, vol. 2, pp. 147- 156, 1959.

[7] C. Berrou, A. Glavieux et P. Thitimajshima, ”Near shannon limit errorcorrecting coding and decoding :turbo-codes," IEEE International Conference on Communications, vol. 2, May 1993.

[8] MacKay et Neal, ”Near Shannon limit performance of low density parity check codes,” Electronics letters, 1997.

[9] M. Luby, M. Mitzenmacher, A. Shokrollahi et D. Spielman, ”Analysis of low density codes and improved designs using irregular graphs," Proceeding of 30th ACM Symp. on Theory of Computing, 1998.

[10] R. G. Gallager, “Low Density Parity Check Codes”, Cambridge, USA: MIT Press, July 1963.

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

[12] H. Xiao-Yu, M. Fossorier et E. ELEFTHERIOU, ”On the computation of the minimum distance of low-density parity-check codes," IEEE International Conference on Communications, 2004.

[13] L. Zhang, Q. Huang, S. Lin, K. Abdel-Ghaffar et I. F. Blake, ”Quasi-cyclic LDPC codes: an algebraic construction rank analysis and codes on latin squares," IEEE Trans. Commun., vol. 58, pp. 3126-3139, Nov 2010.

[14] F. Hamidi-Sepehr, A. Nimbalker and G. Ermolaev, "Analysis of 5G LDPC Codes Rate-Matching Design," 2018 IEEE 87th Vehicular Technology Conference (VTC Spring), 2018, pp. 1-5.

[15] J.H. Bae, A.A. Abotabl, H. Lin, K. Song and J. Lee “An overview of channel coding for 5G NR cellular communications, " APSIPA Transactions on Signal and Information Processing, vol.8, p. e17 2019.

[16] H. Li, B. Bai, X. Mu, J. Zhang and H. Xu, "Algebra-Assisted Construction of Quasi-Cyclic LDPC Codes for 5G New Radio," in IEEE Access, vol. 6, pp. 50229-50244, 2018.

[17] M. Xu, J.Wu and M. Zhang, "A modified Offset Min-Sum decoding algorithm for LDPC codes," 2010 3rd International Conference on Computer Science and Information Technology, 2010, pp. 19-22.

[18] S. Ahmadi, “ 5G NR, Architecture, Technology, Implementation, and Operation of 3GPP New Radio Standard”,London, United Kingdom : Academic Press, an imprint of Elsevier, 2019.

[19] S. Shao et al., "Survey of Turbo, LDPC, and Polar Decoder ASIC Implementations," in IEEE Communications Surveys and Tutorials, vol. 21, no. 3, pp. 2309-2333, thirdquarter 2019.

[20] T. Srisupha, K. Mueadkhunthod, W. Phakphisut, S. Khittiwitchayakul, K. Puntsri and T. Sopon, "Development of 5G LDPC Experimental Kit," 2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), Phuket, Thailand, 2022, pp. 856-859.

[21] S. Belhadj and M. L. Abdelmounaim, "On error correction performance of LDPC and Polar codes for the 5G Machine Type Communications," 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 2021, pp. 1-4.

[22] K. -Y. Li, H. -Y. Li, L. Deng, X. -H. Sun, Y. Wang and Z. -P. Shi, "Improved Informed Dynamic Scheduling Strategies for 5G LDPC Codes," 2023 IEEE 23rd International Conference on Communication Technology (ICCT), Wuxi, China, 2023, pp. 74-78.

[23] K. Tian and H. Wang, "A Novel Base Graph Based Static Scheduling Scheme for Layered Decoding of 5G LDPC Codes," in IEEE Communications Letters, vol. 26, no. 7, pp. 1450-1453, July 2022.

[24] K. D. Rao and T. A. Babu, "Performance Analysis of QC-LDPC and Polar Codes for eMBB in 5G Systems," 2019 International Conference on Electrical, Electronics and Computer Engineering (UPCON), Aligarh, India, 2019, pp. 1-6.

[25] H. Wu and H. Wang, "A High Throughput Implementation of QC-LDPC Codes for 5G NR," in IEEE Access, vol. 7, pp. 185373-185384, 2019.

[26] K. Sun and M. Jiang, "A Hybrid Decoding Algorithm for Low-Rate LDPC codes in 5G," 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP), 2018, pp. 1-5.

[27] F. Ivanov and A. Kuvshinov, "On the Serial Concatenation of LDPC Codes," 2023 16th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS), Nis, Serbia, 2023, pp. 228-231.

[28] L. Song, S. Yu and Q. Huang, "Low-density parity-check codes: Highway to channel capacity," in China Communications, vol. 20, no. 2, pp. 235-256, Feb. 2023

[29] S. Liao, Y. Zhan, Z. Shi and L. Yang, "A High Throughput and Flexible Rate 5G NR LDPC Encoder on a Single GPU," 2021 23rd International Conference on Advanced Communication Technology (ICACT), PyeongChang, Korea (South), 2021, pp. 29-34.

[30] C. Tarver, M. Tonnemacher, H. Chen, J. Zhang and J. R. Cavallaro, "GPU-Based, LDPC Decoding for 5G and Beyond," in IEEE Open Journal of Circuits and Systems, vol. 2, pp. 278-290, 2021.

[31] J. Nadal and A. Baghdadi, "Parallel and Flexible 5G LDPC Decoder Architecture Targeting FPGA," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 29, no. 6, pp. 1141-1151, June 2021.

[32] Y. Xu, W. Wang, Z. Xu and X. Gao, "AVX-512 Based Software Decoding for 5G LDPC Codes," 2019 IEEE International Workshop on Signal Processing Systems (SiPS), Nanjing, China, 2019, pp. 54-59.

[33] Mody Sy, "Optimization Strategies for Low Latency 5G NR LDPC Decoding on General Purpose Processor", 5-th IEEE International Conference on Control, Communication and Computing (ICCC 2023),Thiruvananthapuram, India, May 19-20, 2023.

[34] 2. M. Li, V. Derudder, K. Bertrand, C. Desset and A. Bourdoux, "High-Speed LDPC Decoders Towards 1 Tb/s," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 68, no. 5, pp. 2224-2233, May 2021.

[35] A. Aronov, L. Kazakevich, J. Mack, F. Schreider and S. Newton, "5G NR LDPC Decoding Performance Comparison between GPU and FPGA Platforms," 2019 IEEE Long Island Systems, Applications and Technology Conference (LISAT), Farmingdale, NY, USA, 2019, pp. 1-6.

[36] S. -K. Ahn, K. -J. Kim, S. Myung, S. -I. Park and K. Yang, "Comparison of Low-Density Parity-Check Codes in ATSC 3.0 and 5G Standards," in IEEE Transactions on Broadcasting, vol. 65, no. 3, pp. 489-495, Sept. 2019.

[37] H. Cui, F. Ghaffari, K. Le, D. Declercq, J. Lin and Z. Wang, "Design of High-Performance and Area-Efficient Decoder for 5G LDPC Codes," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 68, no. 2, pp. 879-891, Feb. 2021.

[38] K. Le Trung, F. Ghaffari and D. Declercq, "An Adaptation of Min-Sum Decoder for 5G Low-Density Parity-Check Codes," 2019 IEEE International Symposium on Circuits and Systems (ISCAS), Sapporo, Japan, 2019.

[39] B. A. Jayawickrama and Y. He, "Improved Layered Normalized Min-Sum Algorithm for 5G NR LDPC," in IEEE Wireless Communications Letters, vol. 11, no. 9, pp. 2015-2018, Sept. 2022.

[40] L. Li, J. Xu, J. Xu and L. Hu, "LDPC design for 5G NR URLLC and mMTC," 2020 International Wireless Communications and Mobile Computing (IWCMC), Limassol, Cyprus, 2020, pp. 1071-1076.

[41] H. Wu and H. Wang, "Decoding Latency of LDPC Codes in 5G NR," 2019 29th International Telecommunication Networks and Applications Conference (ITNAC), Auckland, New Zealand, 2019, pp. 1-5.

[42] J. Dai et al., "Learning to Decode Protograph LDPC Codes," in IEEE Journal on Selected Areas in Communications, vol. 39, no. 7, pp. 1983-1999, July 2021.

[43] N. Shah and Y. Vasavada, "Neural Layered Decoding of 5G LDPC Codes," in IEEE Communications Letters, vol. 25, no. 11, pp. 3590-3593, Nov. 2021.

[44] Y. Tang, L. Zhou, S. Zhang, C. Chen and L. Wang, "Normalized Neural Network for Belief Propagation LDPC Decoding," 2021 IEEE International Conference on Networking, Sensing and Control (ICNSC), Xiamen, China, 2021, pp. 1-5.

[45] K. Andreev, A. Frolov, G. Svistunov, K. Wu and J. Liang, "Deep Neural Network Based Decoding of Short 5G LDPC Codes," 2021 XVII International Symposium "Problems of Redundancy in Information and Control Systems" (REDUNDANCY), Moscow, Russian Federation, 2021, pp. 155-160

[46] Jinghu Chen, R. M. Tanner, C. Jones and Yan Li, "Improved min-sum decoding algorithms for irregular LDPC codes," Proceedings. International Symposium on Information Theory, 2005. ISIT 2005., Adelaide, SA, Australia, 2005, pp. 449-453,

[47] 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.

[48] 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.

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

[50] 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.

[51] A. Elkelesh, M. Ebada, “Polar Codes", Institut fur Nachrichtenubertragung. WebDemos.

[52] 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.

[53] 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.

[54] 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.

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

[56] 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.

[57] 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.

[58] U. U. Fayyaz and J. R. Barry, "Low-Complexity Soft-Output Decoding of Polar Codes," in IEEE Journal on Selected Areas in Communications, vol. 32, no. 5, pp. 958-966, May 2014

[59] B. Yuan and K. K. Parhi, "Early Stopping Criteria for Energy-Efficient Low-Latency Belief-Propagation Polar Code Decoders," in IEEE Transactions on Signal Processing, vol. 62, no. 24, pp. 6496-6506, Dec.15, 2014.

[60] V. Bioglio, C. Condo and I. Land, "Design of Polar Codes in 5G New Radio," in IEEE Communications Surveys and Tutorials, vol. 23, no. 1, pp. 29-40, Firstquarter 2021.

[61] H. Gamage, V. Ranasinghe, N. Rajatheva and M. Latva-aho, "Low Latency Decoder for Short Blocklength Polar Codes," 2020 European Conference on Networks and Communications (EuCNC), Dubrovnik, Croatia, 2020, pp. 305-310.

[62] M. Geiselhart, A. Elkelesh, M. Ebada, S. Cammerer and S. ten Brink, "CRC-Aided Belief Propagation List Decoding of Polar Codes," 2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 2020, pp. 395-400, doi: 10.1109/ISIT44484.2020.9174249.

[63] J. Piao, J. Dai and K. Niu, "CRC-Aided Sphere Decoding for Short Polar Codes," in IEEE Communications Letters, vol. 23, no. 2, pp. 210-213, Feb. 2019.

[64] Y. Shen, W. Zhou, Y. Huang, Z. Zhang, X. You and C. Zhang, "Fast Iterative Soft-Output List Decoding of Polar Codes," in IEEE Transactions on Signal Processing, vol. 70, pp. 1361-1376, 2022

[65] A. Cavatassi, T. Tonnellier and W. J. Gross, "Asymmetric Construction of Low-Latency and Length-Flexible Polar Codes," ICC 2019 - 2019 IEEE International Conference on Communications (ICC), Shanghai, China, 2019, pp. 1-6.

[66] Z. B. Kaykac Egilmez, L. Xiang, R. G. Maunder and L. Hanzo, "The Development, Operation and Performance of the 5G Polar Codes," in IEEE Communications Surveys and Tutorials, vol. 22, no. 1, pp. 96-122, Firstquarter 2020.

[67] F. Ercan, T. Tonnellier, N. Doan and W. J. Gross, "Practical Dynamic SC-Flip Polar Decoders: Algorithm and Implementation," in IEEE Transactions on Signal Processing, vol. 68, pp. 5441-5456, 2020.

[68] C. Kestel et al., "A 506Gbit/s Polar Successive Cancellation List Decoder with CRC," 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, UK, 2020.

[69] A. Ç. Arli and O. Gazi, "A survey on belief propagation decoding of polar codes," in China Communications, vol. 18, no. 8, pp. 133-168, Aug. 2021.

[70] M. Zhu, M. Jiang and C. Zhao, "Adaptive Belief Propagation Decoding of CRC Concatenated NR LDPC and Polar Codes," in IEEE Transactions on Communications, vol. 70, no. 8, pp. 4991-5003, Aug. 2022.

[71] C. Sun, Z. Fei, C. Cao, X. Wang and D. Jia, "Low Complexity Polar Decoder for 5G Embb Control Channel," in IEEE Access, vol. 7, pp. 50710-50717, 2019.

[72] Y. Wang, S. Zhang, C. Zhang, X. Chen and S. Xu, "A Low-Complexity Belief Propagation Based Decoding Scheme for Polar Codes - Decodability Detection and Early Stopping Prediction," in IEEE Access, vol. 7, pp. 159808-159820, 2019.

[73] L. Xiang, S. Zhong, R. G. Maunder and L. Hanzo, "Reduced-Complexity Low-Latency Logarithmic Successive Cancellation Stack Polar Decoding for 5G New Radio and Its Software Implementation," in IEEE Transactions on Vehicular Technology, vol. 69, no. 11, pp. 12449-12458, Nov. 2020.

[74] Z. Liu, R. Liu and H. Zhang, "High-Throughput Adaptive List Decoding Architecture for Polar Codes on GPU," in IEEE Transactions on Signal Processing, vol. 70, pp. 878-889, 2022.

[75] H. Rezaei, V. Ranasinghe, N. Rajatheva, M. Latva-aho, G. Park and O. -S. Park, "Implementation of Ultra-Fast Polar Decoders," 2022 IEEE International Conference on Communications Workshops (ICC Workshops), Seoul, Korea, Republic of, 2022, pp. 235-241.

[76] G. Sarkis, P. Giard, C. Thibeault and W. J. Gross, "Autogenerating software polar decoders," 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Atlanta, GA, USA, 2014, pp. 6-10.

[77] G. Sarkis, P. Giard, A. Vardy, C. Thibeault and W. J. Gross, "Increasing the speed of polar list decoders," 2014 IEEE Workshop on Signal Processing Systems (SiPS), Belfast, UK, 2014, pp. 1-6.

[78] D. Kam, H. Yoo and Y. Lee, "Ultralow-Latency Successive Cancellation Polar Decoding Architecture Using Tree-Level Parallelism," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 29, no. 6, pp. 1083-1094, June 2021.

[79] N. Doan, S. A. Hashemi, E. N. Mambou, T. Tonnellier and W. J. Gross, "Neural Belief Propagation Decoding of CRC-Polar Concatenated Codes," ICC 2019 - 2019 IEEE International Conference on Communications (ICC), Shanghai, China, 2019, pp. 1-6.

[80] S. A. Hashemi, N. Doan, T. Tonnellier and W. J. Gross, "Deep-Learning-Aided Successive-Cancellation Decoding of Polar Codes," 2019 53rd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2019, pp. 532-536.

[81] N. Doan, S. A. Hashemi and W. J. Gross, "Decoding Polar Codes with Reinforcement Learning," GLOBECOM 2020 - 2020 IEEE Global Communications Conference, Taipei, Taiwan, 2020.

[82] N. Doan, S. A. Hashemi, F. Ercan and W. J. Gross, "Fast SC-Flip Decoding of Polar Codes with Reinforcement Learning," ICC 2021 - IEEE International Conference on Communications, Montreal, QC, Canada, 2021, pp. 1-6.

[83] I. Wodiany and A. Pop, "Low-Precision Neural Network Decoding of Polar Codes," 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cannes, France, 2019.

[84] C. Wen, J. Xiong, L. Gui and L. Zhang, "A BP-NN Decoding Algorithm for Polar Codes," 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP), Xi'an, China, 2019.

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

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

[87] 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.

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

[89] 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.

[90] 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.

[91] 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.

[92] Z. Xie, P. Chen, R. Chen and Y. Fang, "Polar Coded Modulation Operated With Physical Network Coding," in IEEE Signal Processing Letters, vol. 28, pp. 997-1001, 2021.

[93] Y. Wang, W. Liu and L. Fang, "Adaptive Modulation and Coding Technology in 5G System," in 2020 International Wireless Communications and Mobile Computing (IWCMC), Limassol, Cyprus, 2020, pp. 159-164.

[94] H. Mu, Z. Ma, G. K. Karagiannidis and P. D. Diamantoulakis, "Bit-interleaved polar coded modulation with iterative successive cancellation list decoding," in China Communications, vol. 19, no. 5, pp. 54-68, May 2022.

[95] K. Tian, R. Liu and R. Wang, "Joint Successive Cancellation Decoding for Bit-Interleaved Polar Coded Modulation," in IEEE Communications Letters, vol. 20, no. 2, pp. 224-227, Feb. 2016.

[96] P. Yuan, M. C. Coşkun and G. Kramer, "Polar-Coded Non-Coherent Communication," in IEEE Communications Letters, vol. 25, no. 6, pp. 1786-1790, June 2021.

[97] H. Mahdavifar, M. El-Khamy, J. Lee and I. Kang, "Polar Coding for Bit-Interleaved Coded Modulation," in IEEE Transactions on Vehicular Technology, vol. 65, no. 5, pp. 3115-3127, May 2016.

[98] S. -L. Shieh, Y. -C. Huang, P. -N. Chen and Y. -M. Li, "Systematic Polar Coded Modulation for Informed Receivers," in IEEE Transactions on Communications, vol. 69, no. 10, pp. 6469-6484, Oct. 2021.

[99] H. Wan and A. Nosratinia, "Short-Block Length Polar-Coded Modulation for the Relay Channel," in IEEE Transactions on Communications, vol. 71, no. 1, pp. 26-39, Jan. 2023.

[100] J. Dai, K. Niu, Z. Si and D. Zhang, "Polar-Coded Spatial Modulation," in IEEE Transactions on Signal Processing, vol. 69, pp. 2203-2217, 2021.

[101] S. Li, M. Cai, L. Jin, Y. Sun, H. Wu and P. Wang, "An Ultra-Reliable Low-Latency Non-Binary Polar Coded SCMA Scheme," in IEEE Transactions on Vehicular Technology, vol. 71, no. 6, pp. 6518-6533, June 2022.

[102] R. Yuan, B. Bai, J. Fang, R. Xu and J. Wang, "Two-Level LDPC-Coded Modulation With Probabilistic Shaping," in IEEE Communications Letters, vol. 27, no. 8, pp. 2019-2023, Aug. 2023.

[103] J. Choi, "MIMO-BICM Iterative Receiver With the EM Based Channel Estimation and Simplified MMSE Combining With Soft Cancellation," in IEEE Transactions on Signal Processing, vol. 54, no. 8, pp. 3247-3251, Aug. 2006.

[104] M. -C. Chiu, "Analysis and Design of Polar-Coded Modulation," in IEEE Transactions on Communications, vol. 70, no. 3, pp. 1508-1521, March 2022.

[105] Y. Xu et al., "Enhancements on Coding and Modulation Schemes for LTE-Based 5G Terrestrial Broadcast System," in IEEE Transactions on Broadcasting, vol. 66, no. 2, pp. 481-489, June 2020.

[106] C. Chen, L. Li, L. Wang, S. Wang, X. Li and G. K. Karagiannidis, "Noncoherent Detection With Polar Codes," in IEEE Access, vol. 7, pp. 6362-6372, 2019.

[107] 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.

[108] 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.

[109] G. Li, Y. Xu, Y. Xu, Z. Huang and L. Jinzhao, "Soft Information Learning of BICM-ID System Based on Deep Learning," 2021 International Wireless Communications and Mobile Computing (IWCMC), Harbin City, China, 2021, pp. 1965-1970.

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

[111] 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.

[112] 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.

[113] 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.

[114] 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.

[115] M. Shirvanimoghaddam et al., "Short Block-Length Codes for Ultra-Reliable Low Latency Communications," in IEEE Communications Magazine, vol. 57, no. 2, pp. 130-137, February 2019.

[116] C. Yue M. Shirvanimoghaddam, B. Vucetic and Y. Li, "Channel Coding and Decoding Schemes for URLLC," in Ultra-Reliable and Low-Latency Communications (URLLC) Theory and Practice: Advances in 5G and Beyond , Wiley, 2023, pp.119-168.

[117] 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.

[118] Yuheng Huang and J. A. Ritcey, "Joint iterative channel estimation and decoding for bit-interleaved coded modulation over correlated fading channels," in IEEE Transactions on Wireless Communications, vol. 4, no. 5, pp. 2549-2558, Sept. 2005.

[119] X. Wu, Y. Wang and C. Li, "Low-Complexity CRC Aided Joint Iterative Detection and SCL Decoding Receiver of Polar Coded SCMA System," in IEEE Access, vol. 8, pp. 220108-220120, 2020.

[120] W. -C. Sun, Y. -C. Su, Y. -L. Ueng and C. -H. Yang, "An LDPC-Coded SCMA Receiver With Multi-User Iterative Detection and Decoding," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 66, no. 9, pp. 3571-3584, Sept. 2019.

[121] 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.

[122] M. Marey and O. A. Dobre, "Iterative Receiver Design for Uplink OFDMA Cooperative Systems," in IEEE Transactions on Broadcasting, vol. 62, no. 4, pp. 936-947, Dec. 2016.

[123] Z. Yuan, F. Liu, W. Yuan, Q. Guo, Z. Wang and J. Yuan, "Iterative Detection for Orthogonal Time Frequency Space Modulation With Unitary Approximate Message Passing," in IEEE Transactions on Wireless Communications, vol. 21, no. 2, pp. 714-725, Feb. 2022.

[124] M. Li, C. Abdel Nour, C. Jégo, J. Yang and C. Douillard, "Efficient iterative receiver for bit-Interleaved Coded Modulation according to the DVB-T2 standard," 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 2011, pp. 3168-3171.

[125] C. Qian, R. Gu, W. Xu, J. Xu and X. You, "Enhancing Wideband Multiuser MIMO Uplink Using Superimposed Pilots: Joint Receiver Design," in IEEE Wireless Communications Letters, vol. 13, no. 4, pp. 1138-1142, April 2024.

[126] M. Xhemrishi, M. C. Coşkun, G. Liva, J. Östman, 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.

[127] 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.

[128] 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.

[129] 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.

[130] 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.

[131] 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.

[132] 3GPP TS 38.212 V16.2.0, “ Technical Specification Group Radio Access Network, Multiplexing and channel coding”, July 2020.

[133] 3GPP TS 38.212 V16.2.0, “ Technical Specification Group Radio Access Network, Services provided by the physical layer”, November 2020.

[134] 3GPP TS 38.211 V16.5.0, “ Technical Specification Group Radio Access Network, Physical Channels and Modulation”, May 2021.

[135] 3GPP TR 38 901 v16.1.0,“ Technical Report, Study on channel model for frequencies from 0.5 to 100 GHz”, Nov 2020.

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

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

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

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

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

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

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

[143] B. Classon, K. Blankenship and V. Desai, "Channel codng for 4G systems with adaptive modulation and codng," in IEEE Wireless Communications, vol. 9, no. 2, pp. 8-13, April 2002.

[144] L. Szczecinski, A. Alvarado and R. Feick, "Closed-form approximation of Coded BER in QAM-based BICM Faded Transmission," 2008 IEEE Sarnoff Symposium, Princeton, NJ, USA, pp. 1-5, 2008.

[145] Stephen B. Wicker,“Error Control Systems for Digital Communication and Storage”, Prentice Hall, July 1994

[146] 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.

[147] 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.

[148] 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.

[149] 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.

[150] M. Lian, C. Hager 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.

[151] H. Ji, S. Park and B. Shim, "Sparse Vector Coding for Ultra Reliable and Low Latency Communications," in IEEE Transactions on Wireless Communications, vol. 17, no. 10, pp. 6693-6706, Oct. 2018.

[152] W. Kim, S. K. Bandari and B. Shim, "Enhanced Sparse Vector Coding for Ultra-Reliable and Low Latency Communications," in IEEE Transactions on Vehicular Technology, vol. 69, no. 5, pp. 5698-5702, May 2020.

[153] X. Zhang, D. Zhang, B. Shim, G. Han, D. Zhang and T. Sato, "Sparse Superimposed Coding for Short-Packet URLLC," in IEEE Internet of Things Journal, vol. 9, no. 7, pp. 5275-5289, 1 April1, 2022.

[154] L. Yang and P. Fan, "Improved Sparse Vector Code Based on Optimized Spreading Matrix for Short-Packet in URLLC," in IEEE Wireless Communications Letters, vol. 12, no. 4, pp. 728-732, April 2023.

[155] Y. Zhang et al., "Sparse Superimposed Vector Transmission for Short-Packet High-Mobility Communication," in IEEE Wireless Communications Letters, vol. 12, no. 11, pp. 1961-1965, Nov. 2023.

[156] Y. Zhang, X. Zhu, Y. Liu, H. Liang, Y. L. Guan and V. K. N. Lau, "Block Sparse Vector Coding based Ultra-Reliable and Low-Latency Short-Packet Transmission," 2024 IEEE Global Communications Conference, Cape Town, South Africa, 2024.

[157] Y. C. Pati, R. Rezaiifar and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 1993.

[158] D. Needel and J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” Appl. Comput. Harmon. Anal., vol. 26, no. 3, pp. 301–321, 2009.

[159] R. Tibshirani, “Regression Shrinkage and Selection Via the Lasso“, Journal of the Royal Statistical Society: Series B (Methodological), Volume 58, Issue 1, January 1996

[160] René Adad, “Principales propriétés des coefficients binomiaux”, on Math-OS https://math-os.com/ptes-coeffts-binomiaux/, October 18, 2017 (accessed on September 23, 2025).

[161] 3GPP TS 38.321 V16.0.0, “Technical Specification Group Radio Access Network, Study on Non-Orthogonal Multiple Access (NOMA) for NR (Release 16)", Dec. 2018.