ITERATIVE DETECTION ALGORITHMS FOR HIGH SPECTRAL EFFICIENCY WIRELESS COMMUNICATION SYSTEMS

by
Qiaolin Shi

Dissertation submitted in fulfilment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

under the supervision of

Dr. Diep N. Nguyen
Prof. Xiaojing Huang

School of Electrical and Data Engineering
University of Technology Sydney
Sydney, Australia

November 2020

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, QIAOLIN SHI, declare that this thesis is submitted in fulfilment of the requirements for the award of DOCTOR OF PHILOSOPHY in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree at any other academic institution except as fully acknowledged within the text. This thesis is the result of a Collaborative Doctoral Research Degree program with Beijing Institute of Technology.

This research is supported by the Australian Government Research Training Program.

Name of Student: Qiaolin Shi

Signature of Student: Production Note:
Signature removed prior to publication.

Date: 21/03/2021

ABSTRACT

With the ever-growing of the number of devices and new applications (e.g., in industry automation, intelligent transportation systems, healthcare) and given the severe bandwidth congestion observed at the sub-6GHz frequency bands, it is critical to develop high spectral efficiency transmission techniques. However, the detection of new transmission signalling becomes more challenging compared to that of its conventional counterpart. In this thesis, advanced iterative detection algorithms for index modulation-aided and faster-than-Nyquist (FTN) signalling-aided communication systems are investigated. First, factor graph-based message passing algorithms are proposed for the joint phase noise (PHN) estimation and signal detection in orthogonal frequency division multiplexing with index modulation (OFDM-IM) systems. The solutions are superior to conventional extended Kalman filter and variational approaches in terms of the robustness to severe PHN, as well as the realistic imperfect channel state information and residual carrier frequency offset. Second, amalgamated belief propagation and mean field message passing methods based iterative detection algorithms are developed for satellite communication systems relying on the dual mode-aided index modulation (Sat-DMIM) over nonlinear dispersive satellite channels. The computational complexity of the proposed detector is further reduced by approximating the nonlinear messages using a Taylor series expansion technique. The bit error rate performance of Sat-DMIM is improved compared with the conventional linear equalizer which directly linearizes the

nonlinear system model. Third, frequency-domain joint channel estimation and signal detection methods using the variational Bayesian framework for FTN systems over frequency-selective fading channels are designed. Taking into account the structured inter-symbol interference imposed by FTN signalling and dispersive channels, reliable estimates of channel coefficients and FTN symbols are obtained by minimizing the variational free energy without the aid of the cyclic prefix. Simulation results show that the FTN system relying on the proposed iterative detection algorithm can significantly improve the spectral efficiency compared to its conventional Nyquist counterpart.

ACKNOWLEDGMENTS

The accomplishment of my PhD thesis is owed to the contributions and supports of many people. First and foremost, I would like to express my appreciation and thanks to my supervisors Dr. Diep N. Nguyen and Prof. Xiaojing Huang for their tremendous encouragement, guidance, constructive comments and enduring patience. Dr. Nguyen not only taught me a lot in the research journey but also took good care of my life. Dr. Nguyen's kind support and valuable guidance go far beyond this thesis, and I feel greatly honoured to be supervised by him. I also want to thank my co-supervisor Prof. Huang for his kind care, as well as wonderful support and valuable advice for my research work. From them, I have learned to be patient and dedicated to the research work, which will undoubtedly influence my life.

I would like to thank Prof. Nan Wu and Prof. Hua Wang from Beijing Institute of Technology for their guidance in my early graduate years, from whom I have learned how to find latest studies that interest me and think rigorously. I am greatly thankful to Prof. Lajos Hanzo from University of Southampton for kindly inviting me to visit his group for one year and conducting collaborative research works. I also thank Prof. Xiaoli Ma from Georgia Institute of Technology and A/Prof. Qinghua Guo from the University of Wollongong for their precious guidance on doing research and paper writing. Collaborating with them has broadened my horizon and improved my research.

I would like to thank Dr. Peiyuan Qin, Dr. Forest Zhu and Dr. Negin Shariati for their kind help during my PhD study. Furthermore, many

thanks to all the staff from the School of Electrical and Data Engineering, University of Technology Sydney, for various forms of help they gave to me. I would like to thank all my colleagues and friends for their support and company.

Last but most importantly, my deepest gratitude and love to my family. My loving husband, Ruiheng Zhang, always puts up with me and holds my hand during this long journey. He always tells me funny jokes to make me laugh. We row upstream in the stormy winds and we also have a lot of happy water fights. My beloved daughter, Yuxi, your beautiful smile is always the sunshine in my life. My best friend, Qi Zhang, is always there with me. Finally, I would like to express my gratefulness to my parents Shengchun and Conghua for their unconditional love, patience and sacrifice.

Dedicated to My Beloved Husband, Daughter and Parents

Contents

List of Figures

List of Tables

List of Publications

Journal publications

- **Qiaolin Shi**, Nan Wu, Hua Wang, Xiaoli Ma, Lajos Hanzo, "Factor Graph Based Message Passing Algorithms for Joint Phase-Noise Estimation and Decoding in OFDM-IM", *IEEE Transactions on Communications (TCOM)*, 2020, 68(5), 2906 – 2921. (Corresponding to Chapter 2)

- **Qiaolin Shi**, Nan Wu, Diep N. Nguyen, Xiaojing Huang, Hua Wang, Lajos Hanzo, "Low-Complexity Iterative Detection for Dual-Mode-Aided Index Modulation in Nonlinear Dispersive Satellite Channels", to be submitted to *IEEE Transactions on Communications (IEEE TCOM)*. (Corresponding to Chapter 3)

- **Qiaolin Shi**, Nan Wu, Xiaoli Ma, Hua Wang, "Frequency-Domain Joint Channel Estimation and Decoding for Faster-than-Nyquist signaling", IEEE Transactions on Communications", *IEEE Transactions on Communications (IEEE TCOM)*, 2018, 66(2), 781-795. (Corresponding to Chapter 4)

- **Qiaolin Shi**, Nan Wu, Hua Wang, Weijie Yuan, "Joint channel estimation and decoding in the presence of phase noise over time-selective flat-fading channels", in *IET Communications*, 2016, 10(5): 577-585.

Conference publications

- **Qiaolin Shi**, Nan Wu, Hua Wang, Diep N. Nguyen, Xiaojing Huang, "Low-Complexity Message Passing Receiver for OFDM-IM in the Presence of Phase Noise", in *IEEE Global Communications Conference (GLOBECOM)*, Dec 2020, pp. $1-6$. (Corresponding to Chapter 2)

- **Qiaolin Shi**, Nan Wu, Hua Wang, "Joint Channel Estimation and Decoding for FTNS in Frequency-Selective Fading Channels", in *IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. $1-6$. (Corresponding to Chapter 4)

- **Qiaolin Shi**, Desheng Shi, Guibo Wang, Nan Wu, Hua Wang, "Joint channel response, phase noise estimation and decoding in time-selective flat Rayleigh fading channels", in *IEEE International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct 2015, pp. $1-5$.

Other

- Nan Wu, Weijie Yuan, Hua Wang, **Qiaolin Shi**, Jingming Kuang, "Frequency-Domain Iterative Message Passing Receiver for Faster-than-Nyquist Signaling in Doubly Selective Channels", in *IEEE Wireless Communications Letters*, 2016, 5(6): 584-587.

- Shaoang Li, Nan Wu, **Qiaolin Shi**, Qinghua Guo, "FTN Signaling-Aided Space-Time Multi-Mode Index Modulation Systems With a GMP-Based Receiver", in *IEEE Access*, 2019, 7: 162898-162912.

- Ruiheng Zhang, Chengpo Mu, Min Xu, Lixin Xu, **Qiaolin Shi**, "Synthetic IR Image Refinement Using Adversarial Learning With Bidirectional Mappings", in *IEEE Access*, 2019, 7: 153734-153750.

- Weijie Yuan, **Qiaolin Shi**, Nan Wu, Qinghua Guo, Yonghui Li, "Gaussian Message Passing Based Passive Localization in the Presence of Receiver Detection Failures", in *IEEE Vehicular Technology Conference (VTC Spring)*, 2018, pp. $1-5$.

Abbreviations

1G	The First Generation
4G	The Fourth Generation
5G	The Fifth Generation
AMP	Approximate Message Passing
AWGN	Additive White Gaussian Noise
AR	Auto Regressive
BER	Bit Error Rate
BP	Belief Propagation
BPSK	Binary Phase Shift Keying
CFO	Carrier Frequency Offset
CIR	Channel Impulse Response
CP	Cyclic Prefix
CS	Compressed Sensing
CSI	Channel State Information
dB	Decibel

DCT Discrete Cosine Transforms

DFT Discrete Fourier Transform

EM Expectation Maximization

EP Expectation Propagation

EXIT Extrinsic Information Transfer

FBMC Filter Bank Multi Carrier

FFT Fast Fourier Fransform

FTN Faster-than-Nyquist

GAMP Generalized Approximate Message Passing

GFDM Generalized Frequency Division Multiplexing

GMP Gaussian Message Passing

ICI Inter-Carrier Interference

IM Index Modulation

IMUX Input Demultiplexer

IoT Internet of Things

IRC Iterative Residual Check

ISI Inter Symbol Interference

KLD Kullback-Leibler Divergence

LDPC Low Density Check Code

LLR Log Likelihood Ratio

LMMSE Linear Minimum Mean Squared Error

LS Least Square

LTE Long Term Evolution

MAP Maximum *A Posteriori*

MF Mean Field

MIMO Multiple-Input Multiple-Output

ML Maximum Likelihood

MMSE Minimum Mean Squared Error

mm-Wave MillimeterWave

OFDM Orthogonal Frequency Division Multiplexing

OFDM-IM OFDM with Index Modulation

OMUX Output Multiplexer Filter

PA Power Amplifier

pdf probability density function

PHN Phase Noise

pmf probability mass function

PSK Phase Shift Keying

QAM Quadrature Amplitude Modulation

QPSK Quadrature Phase Shift Keying

RSC Recursive Systematic Convolutional

SC-IM Single-Carrier with Index Modulation

SIC Successive Interference Cancellation

SISO Soft-In Soft-Out

SM Spatial Modulation

SNR Signal-to-Noise Ratio

SPA Sum Product Algorithm

VB Variational Bayesian

VMP Variational Message Passing

Notation

Boldface capital (small) letters denote matrices (vectors).

$\mathbf{A}_{:,i}$ denotes the ith column of matrix $\mathbf{A}$.

$\mathbf{A}_{i,:}$ denotes the ith row of matrix $\mathbf{A}$.

$\mathbf{A}_{i,k}$ denotes the i,kth element of matrix $\mathbf{A}$.

$\mathcal{D}(\mathbf{a})$ denotes the diagonal matrix constructed from the vector $\mathbf{a}$.

$\mathcal{D}(\mathbf{A})$ denotes the diagonal matrix with the diagonal elements of square matrix $\mathbf{A}$ on its diagonal.

$E\{\cdot\}$ denotes the expectation operation.

$V\{\cdot\}$ denotes the (co)variance operation.

$\mathrm{tr}(\cdot)$ denotes the trace operator.

$\mathbb{R}$ denotes the field of reals.

$\mathbb{C}$ denotes the field of complex.

$\propto$ denotes proportionality.

$\mathbf{F}$ denotes the unitary DFT matrix.

$\mathbf{\Phi}$ denotes the unitary DCT matrix.

$\mathbf{I}$ denotes the identity matrix.

$\mathbf{0}$ denotes the all-zeros matrix.

$\mathbf{1}$ denotes the all-one column-vector.

$(\cdot)^*$ denotes the conjugate operation.

$(\cdot)^T$ denotes the transpose operation.

$(\cdot)^H$ denotes the Hermitian operation.

$(\cdot)^{-1}$ denotes the inverse operation.

$\star$ denotes the linear convolution.

$\otimes$ denotes the cyclic convolution.

$\odot$ denotes the element-wise product.

$\hat{m}_x$ denotes the mean of the random variable x.

$\hat{v}_x$ denotes the mean of the random variable x.

$p(\cdot)$ denotes the pdf of a continuous random variable.

$P(\cdot)$ denotes the pmf of a discrete random variable.

Chapter 1

Introduction

This chapter provides an overview of this thesis. In Section 1.1, the evolution of wireless communication systems is discussed, addressing the explosive demand for high-speed transmission over limited spectrum resources both in the next-generation terrestrial cellular networks and satellite systems. Then the advanced transmission technologies - index modulation and faster-than-Nyquist signalling are reviewed in Sections 1.2 and 1.3, respectively. In Section 1.4, the objectives and main contributions are summarized followed by the organization of this thesis.

1.1 Background

Since the 1970s when the first cellular concept was developed in Bell Laboratories, Holmdel, NJ, wireless communication has greatly changed the way human live and work. Recently, with the rapid development of the Internet of Things (IoT) and the Internet of Vehicles, the demand for data traffic has dramatically increased [1]. In just a few decades, wireless communication has rapidly evolved from the first generation (1G) to the fourth generation (4G). The fifth-generation (5G) mobile communication system is now gradually commercialized with the aim of increasing user data rate and system capacity, supporting a massive number of connections and providing various emerging services [2]. However, it is extremely difficult for the 5G system alone to achieve all the aforementioned performance goals [3]. The integration between satellite systems and terrestrial wireless systems has been a

new trend in the next generation of mobile cellular networks (6G) [4, 5]. Under the framework of satellite-terrestrial integrated network (STIN), satellite communication (Satcom) will play a key role due to its wide coverage and flexibility [6]. Compared with terrestrial cellular systems, Satcom can provide more economical and stable communication services in remote areas, especially in areas with complex geographical environments such as mountains, jungles, and oceans [7].

The increasing demand for a large number of users for high-speed and high-quality multimedia services has made the frequency bands in the sub 6GHz range more and more crowded and scarce. At the same time, the energy consumption problem imposed by the information and communication technology (ICT) industry is becoming more serious [8]. Therefore, it is of great importance to developing spectral- and energy-efficient communication techniques in order to make full use of the scarce spectrum resources and the precious energy resources in future networks. Non-orthogonal signalling [9] and index modulated signalling [10, 11] are considered as promising solutions to improve both the spectral efficiency and the energy efficiency.

On one hand, the index modulation (IM) technique has attracted great interest due to its flexibility of striking a trade-off among spectral efficiency, energy efficiency and transmission reliability [11]. It conveys additional information by the indices of the transmission medium, such as antennas [12], subcarriers [13] and times slots [14]. However, the detection for index modulated signalling becomes more difficult compared with that for conventional systems [11]. This is due to that the receiver has to jointly detect the index modulated information and conventional constellation symbols while taking full account of the activation pattern constraint [12–14].

On the other hand, non-orthogonal waveforms, including filter-bank multi-carrier (FBMC) transmission, generalized frequency division multiplexing (GFDM), and faster-than-Nyquist (FTN) signalling, are able to transmit more information without consuming more resources compared with conventional orthogonal waveforms [15]. By introducing the non-orthogonality, faster-than-Nyquist signalling [16] has become a promising candidate for future communication systems, due to its ability to send pulses beyond the Nyquist rate while preserving the same signalling bandwidth.

However, FTN signalling also results in inter-symbol interference (ISI) and colored noise at the receiver side [17]. Moreover, the detection of FTN signalling in unknown frequency-selective fading channels becomes more challenging as the bandwidth increases [18].

Exploiting the high-frequency bands is also an effective means to relieve the increasing congestion of the existing communication frequency bands [19]. However, the physical channel conditions become less favorable in extremely high frequency [20]. Firstly, the phase noise (PHN) originated from the local oscillator instabilities of both the transmitter and receiver becomes stronger. This is one of the major impairments to the performance of coherent communication systems [21,22]. Secondly, the power amplifier (PA) is inevitably driven near or at the saturation region for the energy efficiency in high frequency, which can arouse serious nonlinear distortions [23]. Thirdly, the frequency-selective fading channels will introduce more severe ISI due to the enormous bandwidth in extremely high frequency [24]. The detection algorithm design for wireless communication systems under the aforementioned channel conditions becomes more challenging, especially for systems relying on new waveforms, such as the index modulated signalling and the non-orthogonal signalling.

Given the above, this thesis studies advanced message passing-based iterative detection algorithms for IM-aided and FTN signalling-aided wireless terrestrial and satellite communication systems operating in the high-frequency bands.

1.2 Index Modulation

Recently, the IM technique has been widely investigated as a strong candidate for future terrestrial cellular systems, due to its unique advantages in terms of bandwidth and power efficiency [10, 12–14, 25]. In contrast to other popular spectrum-efficient techniques, such as the non-orthogonal transmission technique [4, 16, 26], where signals are designed to be superimposed to share the same spectrum resource or time slots, IM introduces a new dimension for transmitting information. By utilizing the indices of the building blocks, additional information bits in IM are

embedded into the transmitted signals without consuming additional power [10]. From another perspective, the IM scheme is able to transfer the saved energy from IM bits to conventional constellation symbols, leading to an improved bit error ratio (BER) performance compared with the traditional schemes under the same total transmission energy [11].

Specifically, as a promising alternative to conventional multi-input multi-output (MIMO) technology, the IM concept in the space domain, the so-called spatial modulation (SM) technique [27] transmits information bits not only by the conventional phase-shift keying (PSK)/quadrature amplitude modulation (QAM) symbols but also the index of the active antenna. Inspired by SM, the concept of IM has been applied to classic orthogonal frequency division multiplexing (OFDM). OFDM is one of the most widespread multi-carrier transmission techniques designed for satisfying the increasing demand for high-rate wireless communications. Compared to single-carrier systems, OFDM is capable of improving the spectral efficiency, despite its reduced detection complexity, in wideband channels. OFDM is now adopted in the long term evolution (LTE)/LTE-enhanced (LTE-A) systems as the downlink physical layer modulation scheme and its corresponding physical layer interface has already been defined in the IEEE 802.15.3c standard for 5G mm-wave communication systems [28]. By relying on the IM technique, OFDM-IM potentially attains a lower peak-to-average power ratio (PAPR) and improved bit error ratio (BER) performance over its conventional OFDM counterpart [13, 29–31].

To elaborate a little further, in the OFDM-IM scheme, information is conveyed not only by conventional PSK/QAM signal constellations but implicitly also by the specific combinations of the activated subcarrier indices. Fig. 1.1 illustrates the QPSK-modulated OFDM and OFDM-IM where 3 out of 4 available subcarriers are activated for conveying QPSK symbols. It can be observed that the OFDM-IM has sparsity at the same transmission rate as that of classic OFDM. The idle subcarriers in OFDM-IM can alleviate the high PAPR problem of OFDM. However, in contrast to the single-tap frequency-domain equalizer of classic OFDM, a joint PSK/QAM symbol and IM symbol detector is required for OFDM-IM systems. Bearing in mind that the values of frequency-domain symbols can be zero or non-zero, a reduced-

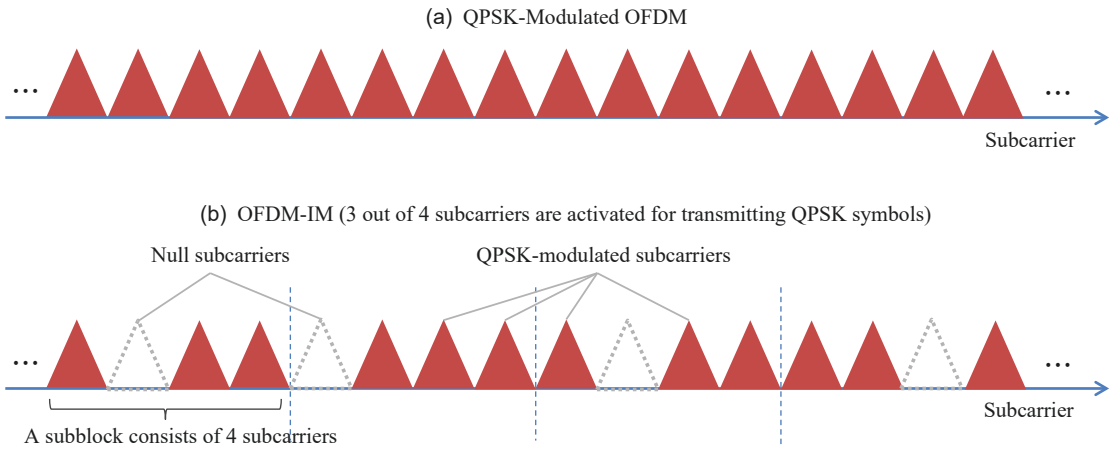

Figure 1.1: Comparison of (a) OFDM and (b) OFDM-IM.

complexity log-likelihood ratio (LLR) detector was proposed in [13] for reducing the excessive complexity of the joint maximum likelihood (ML) receiver. Inspired by the benefits of the greedy pursuit algorithms conceived for the estimation of sparse vectors, a low-complexity iterative residual check (IRC)-based detector is developed in [32], where the associated hard decisions are refined by making locally optimal choices at each step.

Nonetheless, as for all emerging techniques, there are also some open issues such as the design of powerful OFDM-IM detectors for realistic channel conditions, because, in practice, OFDM-IM is extremely sensitive both to dispersive time-varying fading channels, as well as to carrier frequency offset (CFO) and to PHN [33], just like classic OFDM. However, in most of the previous studies, quasi-static frequency-selective fading channels are considered, where the channel state information (CSI) is assumed to be perfectly known at the OFDM-IM receiver [32, 34–36]. As a further aggravation, in contrast to the channel impulse response (CIR) and CFO, which vary slowly with reference to the OFDM symbol rate, the PHN arising from the voltage-controlled oscillator's jitter fluctuates rapidly and thus cannot be eliminated with the aid of pilots. The existing papers mainly focus on the PHN mitigation techniques of classic OFDM [37–39], whilst PHN estimation has not been considered in OFDM-IM systems. Therefore, efficient PHN estimation techniques have to be designed to attain reliable BER performance and

glean multiple advantages from OFDM-IM schemes. To this end, this thesis will study the joint PHN estimation and OFDM-IM signalling detection algorithms in frequency-selective fading channels.

Compared with the aforementioned IM-aided multi-carrier systems, the broadband single-carrier systems relying on IM (SC-IM) have a lower PAPR and thus are less sensitive to nonlinear distortions. The SC-IM scheme is shown to be more desirable for Satcom and for the IoT devices which have various constraints and limitations, since the complexity of the SC-IM transmitter in IoT devices is much lower than that of OFDM. Note that although various IM techniques have attracted considerable research and industry attention in terrestrial networks, the application of them in Satcom is a new yet challenging topic, and so far no relative work has been published. On the one hand, SM is not suitable for Satcom, because it is impossible to identify the indices of active antennas due to the height of satellites. On the other hand, although both OFDM and SC modulations have been widely employed by Satcom, the receiver design of IM-aided OFDM and SC becomes more challenging in Satcom. In contrast to existing IM-based terrestrial communication systems, which are mainly formulated in linear frequency-flat [12] or linear frequency-selective Rayleigh fading channels [13, 14, 40], Satcom generally suffers from nonlinear and dispersive distortions [41].

Satcom can provide economical access in remote areas and reliable services for emergency communications. With this regard, the next-generation terrestrial wireless communication systems have been suggested to be integrated with the space-based networks for providing seamless global coverage [3, 4, 26, 42, 43]. In order to keep up with the rapid growth of terrestrial communications, Satcom should exploit innovative techniques to improve its spectrum utilization and energy efficiency. Compared with single-mode-aided SC-IM, the dual-mode(DM)-aided SC-IM has a higher spectral efficiency [44]. The DM-aided SC-IM scheme, unlike its single-mode counterpart, conveys symbols drawn from a secondary constellation by inactive transmit entities of original IM. By designing the primary and secondary constellations to ensure they distinguish from each other, the DM-aided SC-IM scheme can achieve a larger minimum Euclidean distance (MED) over its single-

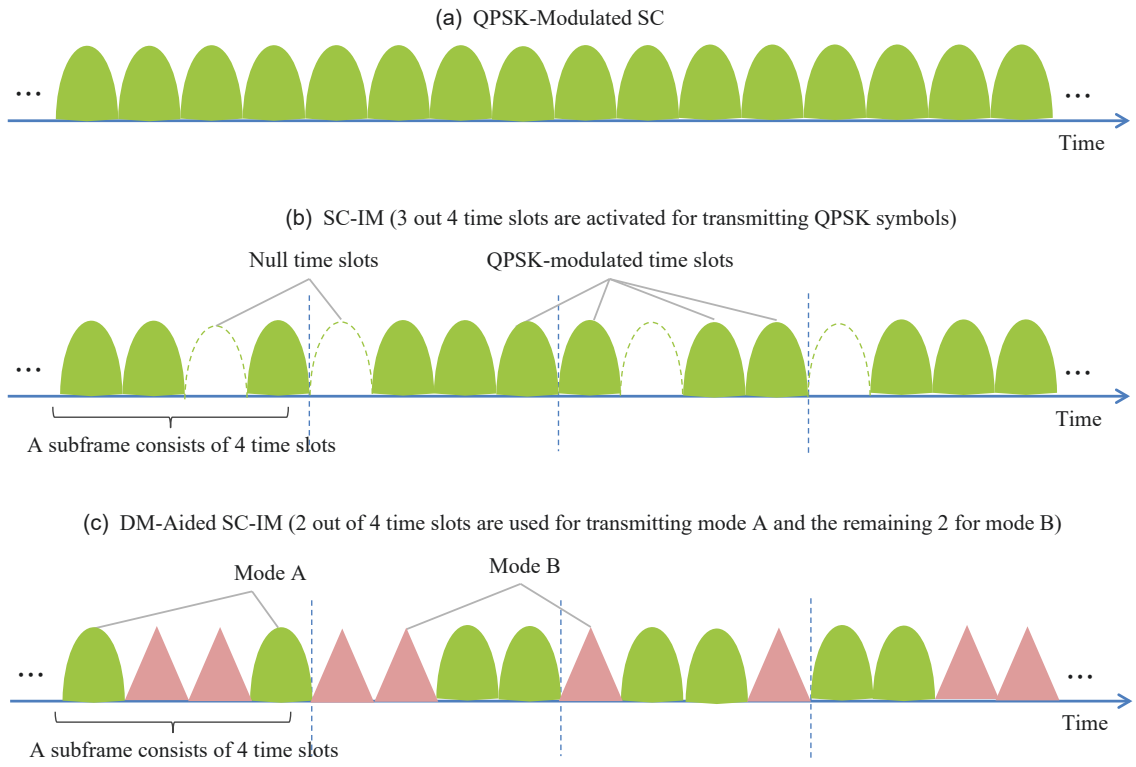

Figure 1.2: Comparison of conventional SC, SC-IM and DM-aided SC-IM.

mode counterpart [45,46]. Fig. 1.2 shows the illustration of conventional SC, SC-IM where 3 out of 4 available time slots are activated for conveying QPSK symbols and dual-mode-aided SC-IM where 2 out of 4 are used for transmitting symbols belong to mode A whilst the remaining 2 for symbols belong to mode B. Here the modes A and B are all QPSK constellation sets and the mode B is obtained by rotating the mode A by a phase of $\pi/4$. Therefore, in DM-aided SC-IM, the information is conveyed both by conventional PSK/QAM symbols and the permutation of the two distinguished modes. It can be observed that the DM-aided SC-IM achieves the highest spectral efficiency amongst the aforementioned three schemes. There exist many null time slots in single-mode-aided SC-IM compared to conventional SC at the same transmission rate.

For DM-aided SC-IM systems, an ML detector was proposed in [44]. By using this detector, the DM-aided SC-IM achieves improved BER performance compared to its conventional single-carrier counterpart. However, the detection for SC-IM systems in existing literatures is predominantly formulated in linear frequency-

selective fading channels. The detection method design for SC-IM in nonlinear channels becomes more challenging. To this end, this thesis will study iterative detection algorithms for single-carrier Satcom relying on DM-aided IM in nonlinear dispersive satellite channels.

1.3 Faster-than-Nyquist signalling

According to the Nyquist criterion, the maximum symbol rate should be $2W$ signals/s for ISI-free transmission over an ideal additive white Gaussian noise (AWGN) channel of bandwidth W. If information symbols are transmitted at a rate higher than that suggested by the Nyquist criterion, ISI is inevitable. In the 1970s, James Mazo investigated the binary sinc-pulse case and accelerated the pulses [47]. Assuming that the ISI-free symbol interval is given by T_0, in the faster-than-Nyquist signalling scheme transmitted symbol's interval T is typically set such that $T = \tau T_0$, where $\tau \leq 1$ is the time accelerate factor. He discovered that the square minimum distance d^2_{min} does not change when the time accelerate factor τ drops below 1 and the pulses become non-orthogonal. This means that around 25% more information bits can be transmitted in the same bandwidth, without deteriorating BER. However, the value of distance d^2_{min} becomes smaller if the transmission rate is further accelerated. Here, the transmission rate at which d^2_{min} first falls below its orthogonal value is the so-called Mazo limit. A similar phenomenon occurs with root raised cosine (rRC) pulses except that the corresponding Mazo limit is also determined by the roll-off factor, i.e., the value of the Mazo limit becomes higher as the roll-off factor increases.

FTN signalling can achieve higher spectral efficiency by sending pulses beyond the Nyquist rate while preserving the same signalling bandwidth [47–50]. It has been considered as a promising modulation format suited for the 5G cellular communications [2, 51, 52]. However, FTN signalling also results in ISI and colored noise at the receiver side. By eliminating the effect of ISI imposed by FTN at the receiver side, improved spectral efficiency can be obtained compared with that of conventional Nyquist systems. Fig. 1.3 illustrates the Nyquist signalling and FTN

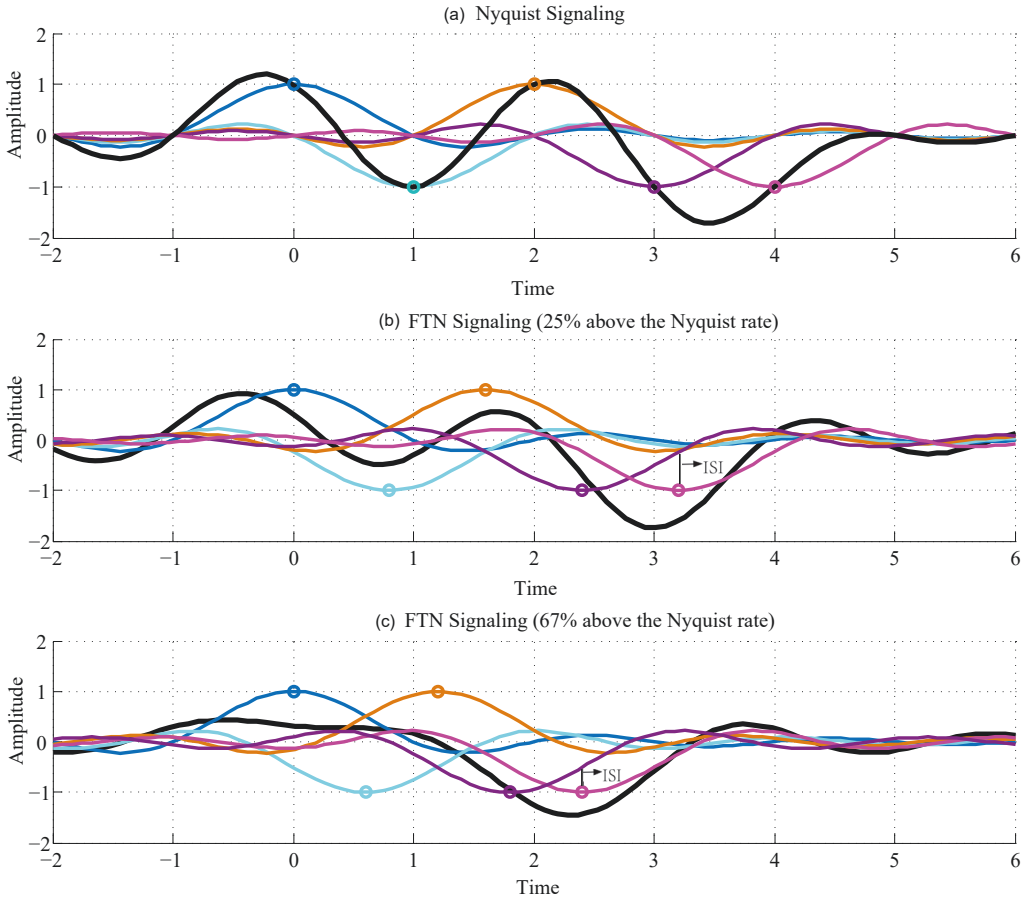

Figure 1.3: Comparison of Nyquist signalling and FTN signalling.

signalling waveforms, where the accelerate factors $\tau = 0.8$ and 0.6 are considered. It can be observed that the spectral efficiency of FTN systems becomes higher as the transmission rate increases. The detection method design, however, becomes more challenging, due to the severer ISI imposed by FTN signalling. The FTN receiver using the maximum likelihood sequence detection (MLSD) method for optimal performance suffers from infeasible computational complexity. Although new silicon technology is making FTN systems a promising alternative, low-complexity detection algorithms are still essential to attain reliable BER performance and glean spectral efficiency advantages from FTN systems [52].

To reduce complexity, the Viterbi algorithm is applied for symbol detection in [53] by taking FTN signalling as a convolutional code. In [54], a maximum *a posteriori* probability (MAP) detector with successive interference cancellation (SIC) is employed. A time-domain M-algorithm Bahl, Cocke, Jelinek, and Raviv

(M-BCJR) equalizer based on a minimum phase model is developed in [55]. The complexity of this algorithm grows linearly to the length of the received symbols but still exponentially to the number of ISI taps and the modulation order. In contrast, [17] leverages cyclic prefix (CP)-based single-carrier frequency-domain equalization (FDE) method in conjunction with a circulant matrix approximation for the ISI effect imposed by FTN signalling to derive a low-complexity receiver based on the minimum mean-square error (MMSE) criteria. An extension is made in [56], where a three-stage-concatenated transceiver structure is provided to reduce the ISI and achieve better BER performance than the two-stage-concatenated FTN signalling counterpart. However, the colored noise at the receiver induced by the faster sampling rate in FTN signalling is not considered in both [17] and [56]. In [57], a colored noise whitening filter based on MMSE weights is designed for a soft-decision FDE-aided FTN receiver. An overlap-based approach is proposed to perform MMSE-FDE in known frequency-selective fading channels in [58]. To compensate for the severe ISI caused by FTN signalling and frequency-selective fading channels, a considerable overlap length of the blocks is required. That dramatically reduces the throughput. In [59], an iterative block decision feedback equalizer (IB-DFE) in the frequency domain is employed to deal with the severe ISI associated with the combined effect of FTN signalling and the time-dispersive channel. In [60], based on two variational inference methods, i.e., mean-field (MF) and Bethe approximations, low-complexity frequency-domain equalizers are proposed for FTN signalling in doubly-selective channels [61].

The existing research on FTN detectors mainly focuses on AWGN channels and fading channels with known CSI at the receiver. However, the perfect CSI is always unknown in practical systems. Therefore, it is necessary to design joint channel estimation and detection in FTN systems. In [18], T. Ishihara et al. proposed a joint frequency-selective fading channel estimation and FTN detection algorithm based on the MMSE criterion. According to the MMSE criterion, the estimates of FTN symbols are calculated in the frequency domain. Then the estimates of CIRs are updated by using a MMSE filtering, exploiting the demodulated FTN symbols as a long pilot sequence in the iterative process. In [62], N. Wu et al. proposed a

hybrid belief propagation (BP)-expectation propagation (EP)-variational message passing (VMP) algorithm in the framework of Forney-style factor graph, where the colored noise introduced by FTN signalling is approximated by an autoregressive process. Considering that the inner product operation between the FTN symbol nodes and the CIR nodes cannot be performed by the BP method, the VMP technique is invoked by introducing an equivalent soft node. Compared with the Nyquist signalling, the transmission rate of the FTN signalling relying on the BP-EP-VMP method is increased by over 40%, without introducing significant BER performance loss. However, the complexity order of this algorithm grows cubically with the number of ISI taps. In [63], by formulating the detection process of FTN signalling in doubly-selective fading channels as a linear state-space model, a low-complexity Gaussian message passing method defined over a factor graph in vector form is developed, where the robustness of the FTN detector to imperfect CSI is considered.

All aforementioned methods approximate the FTN symbols as Gaussian variables, since taking into account the exact discrete distributions of data symbols over the signal constellation leads to intractable complexity. Moreover, all existing frequency-domain detection algorithms require inserting cyclic prefix, which consumes considerable transmission energy and wastes spectrum. To this end, this thesis will study the joint channel estimation and FTN signal detection over frequency-selective fading channels without the aid of cyclic prefix. The bit error rate performance of FTN systems will be improved by taking into account the discrete nature of data symbols.

1.4 Thesis Objectives and Organization

In summary, the main aims of this thesis are as follows:

i The OFDM-IM systems inherit the drawbacks from OFDM, including the vulnerability to time-varying phase noise and carrier frequency offset. Therefore, it is necessary to design low complexity iterative detection algorithms while eliminating the effect of phase noise in the face of realistic imperfect channel

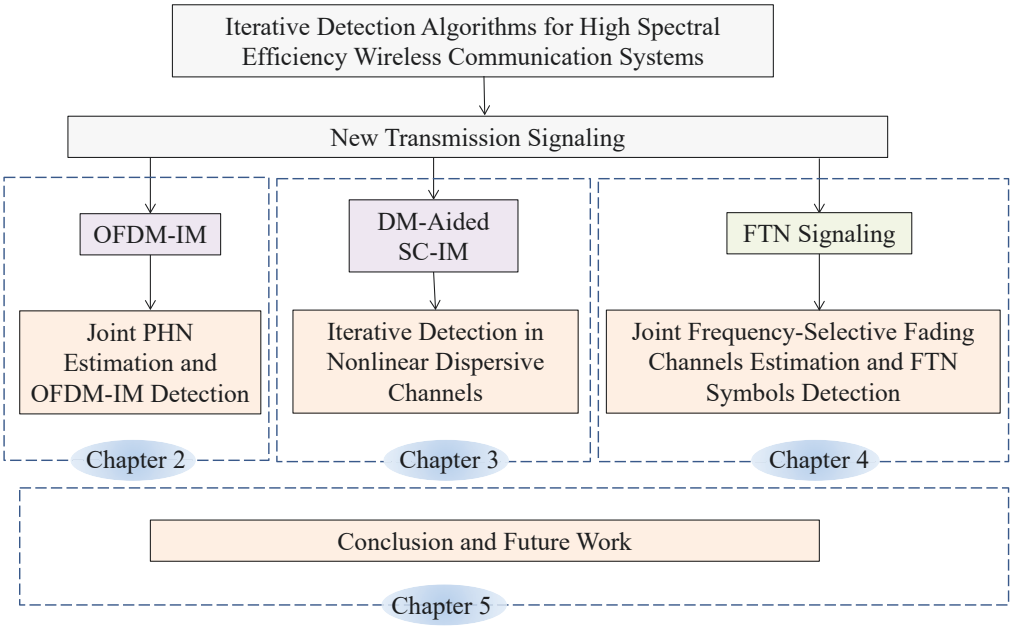

Figure 1.4: Thesis Organization.

state information and residual carrier frequency offset. (Chapter 2).

ii The Sat-DMIM systems generally suffer from nonlinear distortions as the high power amplifiers are often driven to their saturation points for reasons of power efficiency. Therefore, it is necessary to design low-complexity iterative detection algorithms for Sat-DMIM systems to eliminate the effect of both linear and nonlinear inter-symbol interference imposed by satellite channels. (Chapter 3)

iii Conventional frequency-domain equalizers for FTN systems always require a large amount of cyclic prefix, which results in significant power and spectral efficiency loss. Moreover, if the structured ISI introduced by FTN signalling is combined with the interference imposed by dispersive channels, significant performance loss and complexity will be introduced. Therefore, it is necessary to exploit the known ISI imposed by FTN signalling, and the ISI introduced by FTN signalling and channel fading should be separated. (Chapter 4)

The remainder of the thesis is organized as follows, as also illustrated in Fig. 1.4.

- Chapter 2: **Factor graph-based message passing algorithms for joint**

PHN estimation and decoding in the OFDM with index modulation (OFDM-IM) scheme are proposed. The robustness of the OFDM-IM scheme relying on the proposed algorithms is stronger than that of its classic OFDM counterpart under severe PHN. The structured a priori information of the frequency-domain OFDM-IM symbols is explicitly constructed to take into account the sparsity of the OFDM-IM symbols. Then the generalized approximate message passing (GAMP) algorithm is invoked for decoupling the vector estimation process into scalar operations. To circumvent the nonlinear and unknown transform matrix problem in the decoupling step of the GAMP, the MF method is employed at the observation node to simplify the update of messages. The BER performance and computational complexity are compared with existing methods, such as the extended Kalman filter, which demonstrate the effectiveness and efficiency of the proposed methods.

- Chapter 3: **The amalgamated BP and MF methods are proposed for the detection of dual mode-aided index modulated single carrier (SC-IM) symbols in nonlinear dispersive channels. The BER performance is improved compared with its conventional single-carrier systems and single mode-aided SC-IM schemes with the same spectral efficiency.** Taking account of the mode-selection pattern constraint, the merged message passing algorithms are developed for iteratively updating the a posteriori estimates of the data symbols while coping with the nonlinear distortions. By using the first-order Taylor series expansion, part of nonlinear ISI terms are linearized, which leads to a reduction of the number of sufficient statistics to be estimated, leading to a low computational complexity at the receiver side.

- Chapter 4: **Frequency-domain joint channel estimation and decoding methods based on the variational Bayesian framework are proposed for FTN signalling transmission over frequency-selective fading channels. Reliable estimates of channel coefficients and data symbols are obtained.** Without the aid of the cyclic prefix, an alternative

frequency-domain system model is derived to take into account the structured ISI and colored noise imposed by FTN signalling, as well as the unknown ISI introduced by the dispersive channels. By minimizing the variational free energy, the a posteriori probability distributions of the channel coefficients and the FTN symbols are iteratively updated. Furthermore, the GAMP technique is invoked for reducing the performance loss from approximating the discrete symbol variables as Gaussian variables.

- In Chapter 5: Finally, the main contributions of this thesis are summarized in the last chapter and some open issues for future work are discussed.

Chapter 2

Iterative Detection for OFDM-IM Systems in the Presence of PHN

To glean benefits from OFDM-IM in the presence of strong PHN, in this chapter, low-complexity joint PHN estimation and decoding methods are developed under the framework of message passing on a factor graph. Both the Wiener process and the truncated discrete cosine transform (DCT) expansion model are considered for approximating the PHN variation.

2.1 Introduction

The PHN arising from the voltage-controlled oscillator's jitter fluctuates rapidly with reference to the OFDM symbol rate and thus cannot be eliminated with the aid of pilots. Indeed, it is quite a challenge to design a low-complexity OFDM-IM receiver in the presence of strong PHN. The unknown time-varying nonlinear effects imposed by PHN result in a loss of subcarrier orthogonality, which inevitably leads to inter-channel interference (ICI). Accordingly, the system implementation and performance are considerably affected, if the PHN is inadequately compensated. Specifically, if each index modulated symbol vector (IM-symbol) in each OFDM-IM subblock is detected independently, the performance erodes owing to the presence of correlated, i.e., colored, PHN. If however several IM-symbols are jointly detected for mitigating the PHN, the complexity of the receiver becomes prohibitive, even if reduced-search

based detectors are used. Therefore, efficient PHN estimation techniques have to be designed to attain reliable BER performance and glean multiple advantages from OFDM-IM schemes. Moreover, wireless systems tend to rely on iterative channel decoding based on the powerful turbo principle to attain a low BER at SNRs close to channel capacity. However, the existing PHN estimation and OFDM symbol detection problems are predominantly formulated for uncoded systems. Against this background, our goal is to propose low-complexity methods for joint PHN estimation and OFDM-IM symbol detection. Very recently, message passing algorithms defined over factor graphs have gained popularity, including both BP [64], as well as MF approximation based variational methods [65] and GAMP techniques [66]. The BP and MF methods defined over factor graphs are derived from the stationary points of the corresponding variational free energy [67]. By forwarding messages from one node of the factor graph to neighboring nodes according to message updating rules (refer to [64] and [65] for BP and MF rules, respectively), the distance measure between a trial distribution and the original objective function to be estimated can be minimized. GAMP constitutes a state-of-the-art BP approximation, where the message passing between variable nodes and factor nodes is simplified by exploiting the central limit theorem and a second-order Taylor series (refer to [66] for GAMP rules). A structured approximate message passing (AMP) detector is developed for spatial modulation aided systems in [68], which is further generalized in [69], where perfectly known block fading Rayleigh channels are assumed, albeit no joint iterative turbo detection is invoked.

Explicitly, the contributions of this chapter are summarized as follows:

- A GAMP based low-complexity joint PHN estimation and decoding algorithm is proposed for OFDM-IM communicating in frequency-selective fading channels. The structured *a priori* information gleaned from the low-density parity-check (LDPC) soft-in soft-out (SISO) decoder in the turbo receiver [70] is specifically constructed to take into account the sparsity of the OFDM-IM symbols. Given that the consecutive frequency-domain IM symbols are not independent due to the existence of PHN, the GAMP algorithm is invoked for decoupling the vector estimation process into scalar operations. To circumvent

the nonlinear and unknown transform matrix problem in the decoupling step of the GAMP [66], the MF method [65] is employed at the observation node (refer to [64] for details about the observation node defined over factor graphs) for approximating the output function f_{out} defined in GAMP.

- Two PHN models, i.e., the Wiener model [37, 38] and the truncated DCT expansion model [71], are considered for representing the PHN. In contrast to linearizing the system model directly as in [37], the time-domain statistics of PHN are explicitly exploited based on the Wiener model. In contrast to [71], where only hard estimates of the DCT coefficients are given by the expectation-maximization algorithm, the probability distributions of the DCT coefficients are provided. By exploiting the PHN estimates, the accuracy of data detection is gradually improved in an iterative fashion.

- The amalgamated BP-MF algorithm is invoked for joint PHN estimation and decoding as a benchmark. Parallel and sequential message passing aided schedules relying on the DCT model and Wiener model, respectively, are introduced for comparison. The complexity of the parallel scheduling is similar to that of the proposed GAMP based algorithms, since the efficiency of fast Fourier transform (FFT) can be exploited. However, some performance loss is observed due to its low convergence rate compared to both the sequential and to the proposed GAMP based methods. As for the Wiener model, the amalgamated BP-MF relying on sequential message passing outperforms the intrinsically parallel GAMP method, albeit at the expense of an increase of complexity.

- The performance of the proposed algorithms is evaluated and compared to that of classic OFDM for both perfect and realistic imperfect channel conditions. Both subcarrier interleaving and coordinate interleaving techniques are used for further improving the BER performance by increasing the diversity order. Our comparisons both to the conventional pilot based method, to the extended Kalman filter (EKF) of [37] and to the variational inference (VI) approach of [38] are also shown. Our simulation results: 1) confirm the robustness of

the OFDM-IM scheme relying on the proposed algorithms both in the face of severe PHN, as well as realistic imperfect CSI and residual CFO compared to its classic OFDM counterpart; and 2) demonstrate the superiority of the proposed algorithm over existing methods.

The rest of this chapter is organized as follows. Section 2.2 presents the system model of FTN signalling over frequency-selective fading channels. In Sections 2.3 and 2.4, the proposed GAMP-based and BP-MF-based algorithms are derived in details, respectively. The performance results of the proposed algorithms are evaluated via Monte Carlo simulations in Section 2.6. Conclusions are drawn in Section 2.5.

2.2 System Model

Let us consider an LDPC[a] [72] coded OFDM-IM system having N_q data subcarriers in each OFDM-IM block, whilst communicating over an $L_{\rm h}$-tap frequency-selective fading channel. Fig. 2.1 depicts the transceiver schematic of the OFDM-IM system. The data bits $\mathbf{c}_q \triangleq [c_1^{(q)}, \ldots, c_{M_q}^{(q)}]^T$ are generated by encoding M_i information bits $\mathbf{b} \triangleq [b_1, \ldots, b_{M_i}]^T$ using a rate-R_c LDPC coder and then by partitioning the resultant $M_c = M_i/R_c$ bits $\mathbf{c} \triangleq [c_1, \ldots, c_{M_c}]^T$ into an integer number of $Q \triangleq M_c/M_q$ OFDM-IM symbols. These M_q bits in the qth block are then splitted into G subgroups each containing M_g bits, i.e., we have $M_g = M_q/G$. Each subgroup of M_g-bits is mapped to an OFDM-IM subblock of length N_g, where $N_g = N_q/G$. Here, N_q is the size of the inverse fast Fourier transform (IFFT) of Fig. 2.1. In each OFDM-IM subblock, only K_g out of N_g available subcarriers determined by the first M_{g1} bits of the incoming M_g-bit sequence are activated for transmitting S-ary classic PSK/QAM symbols, while the remaining $(N_g - K_g)$ inactive subcarriers are set to zero. In addition to the M_{g1} bits transmitted by the active subcarriers, OFDM-IM also transmits $M_{g2} \triangleq M_g - M_{g1}$ bits by mapping them to K_g S-ary PSK/QAM symbols conveyed by the K_g active subcarriers[b]. Therefore, we have

[a]Unlike LDPC decoding, trellis-based decoding, as, for example, with convolutional codes and turbo codes, involves recursive calculations that prevent parallel computation.

[b]Increase the value of M_{g1} can improve the power efficiency of the OFDM-IM scheme

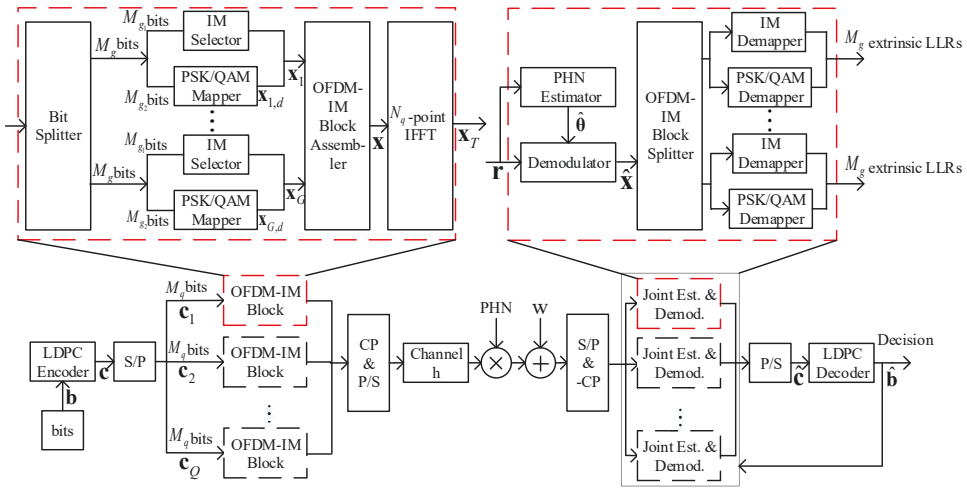

Figure 2.1: The transceiver of the LDPC-coded OFDM-IM system.

$M_g = M_{g1} + M_{g2} = \left\lfloor \log_2 \binom{N_g}{K_g} \right\rfloor + K_g \log_2 S$, where $\binom{N_g}{K_g}$ denotes the binomial coefficient representing the number of subcarrier combinations. Then the M_{g1} bits determine $\Upsilon = 2^{\left\lfloor \log_2 \binom{N_g}{K_g} \right\rfloor}$ legitimate combinations of the active subcarriers. We define the indicator set $\mathbf{\Omega} = \{\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_\Upsilon\}$ associated with $\boldsymbol{\omega}_\gamma = [\omega_{\gamma,1}, \ldots, \omega_{\gamma,N_g}]^T \in \{0,1\}^{N_g}$, where $\omega_{\gamma,i} = 1$ if the ith subcarrier is activated, and $\omega_{\gamma,i} = 0$ otherwise. The coded bits corresponding to $\boldsymbol{\omega}_\gamma$ are represented by $[c_1^{\boldsymbol{\omega}_\gamma}, \ldots, c_{M_{g1}}^{\boldsymbol{\omega}_\gamma}]^T$.

Specifically, we assume that the indices activated in the gth OFDM-IM subblock are given by $\mathcal{I}_g = \{i_{g,1}, \ldots, i_{g,K_g}\}$, $g = 1, \ldots, G$, where $i_{g,k_g} \in [1, \ldots, N_g]$, and the non-zero symbols assigned to the K_g active subcarriers are collected as $\mathbf{x}_{g,d} = [x_{g,d}(1), \ldots, x_{g,d}(K_g)]^T$, $g = 1, \ldots, G$, where $x_{g,d}(k_g) \in \mathcal{S}$ and $\mathcal{S} = \{s_1, \ldots, s_S\}$ denotes the normalized S-PSK/QAM symbol set. The coded bits corresponding to the mth constellation symbol s_m in the set are denoted by $[c_1^{s_m}, \ldots, c_{\log_2 S}^{s_m}]^T$. Then the K_g data symbols are allocated to K_g active subcarriers of the index set in $\mathcal{I}_g$, yielding a subblock of symbols $\mathbf{x}_g$, which can be expressed as $\mathbf{x}_g = \mathbf{I}_g^0 \mathbf{x}_{g,d} \triangleq [x_g(1), \ldots, x_g(N_g)]^T$, $g = 1, \ldots, G$, where $\mathbf{I}_g^0$ consists of the K_g columns from $\mathbf{I}_{N_g}$ having the indices given by $\mathcal{I}_g$. The index modulated symbol vector $\mathbf{x}_g$ in the gth OFDM-IM subblock is referred to as the gth IM-symbol. The activation pattern constraint here is that $\mathbf{x}_g^{\mathrm{AP}} \in \mathbf{\Omega}$, where $x_g^{\mathrm{AP}}(n) = 0$ if $x_g(n) = 0$ and $x_g^{\mathrm{AP}}(n) = 1$ otherwise. In other words, the set of Υ legitimate combinations is used

at both the transmitter and receiver sides, and the transmitter chooses an index combination from the set for each OFDM-IM subblock. After concatenating all symbol vectors from G subblocks by the OFDM-IM block assembler of Fig. 2.1, a main OFDM-IM block can be written as $\mathbf{x} = [\mathbf{x}_1^T, \ldots, \mathbf{x}_G^T]^T \triangleq [x_1, \ldots, x_{N_q}]^T \in \mathbb{C}^{N_q \times 1}$, where the nth element in $\mathbf{x}$ obeys $x_n \in \{\mathcal{S}, 0\}$. Note that with a little abuse of notation, we may denote the bit vector transmitted in the gth subblock by $\mathbf{c}_g = [c_{g,1}^{\mathrm{IM}}, \ldots, c_{g,M_{g1}}^{\mathrm{IM}}, c_{g,1}^{\mathcal{S}}, \ldots, c_{g,M_{g2}}^{\mathcal{S}}]^T \in \{0,1\}^{M_g}$, where $c_{g,i}^{\mathrm{IM}}$, $i = 1, \ldots, M_{g1}$, is the ith subcarrier selection bit, and $c_{g,\jmath}^{\mathcal{S}}$, $\jmath = 1, \ldots, M_{g2}$, the $\jmath$th symbol mapping bit.

By defining the total number of active indices in an OFDM-IM block as $K_q \triangleq GK_g$, we have $E[\mathbf{x}^H\mathbf{x}] = K_q$. At the transmitter, the IFFT having a normalization factor $\frac{N_q}{\sqrt{K_q}}$ is employed for generating the time-domain symbol vector $\mathbf{x}_{\mathrm{T}} = \frac{1}{\sqrt{K_q}}\mathbf{F}^H\mathbf{x}$, where $\mathbf{F} \in \mathbb{C}^{N_q \times N_q}$ is the discrete Fourier transform (DFT) matrix associated with $\mathbf{F}^H\mathbf{F} = N_q\mathbf{I}_{N_q}$. Thus, the transmitted power is normalized, i.e., $E[\mathbf{x}_{\mathrm{T}}^H\mathbf{x}_{\mathrm{T}}] = N_q$. Then, an L_{cp}-length cyclic prefix (CP), which has to be longer than the channel's delay spread for the sake of eliminating the effects of the inter-symbol-interference (ISI) over frequency-selective fading channels, is added at the front of the time-domain symbols. Then the signal generated at the transmitter is sent over an L_{h}-tap frequency-selective fading channel. Initially, the CIRs are assumed to be perfectly known at the receiver, but this assumption will be relaxed in Section V, where the effects of imperfect CIRs will also be considered. In the presence of unknown PHN distortion, the N_q-point complex baseband received signal sequence $\mathbf{r} \in \mathbb{C}^{N_q \times 1}$ of an OFDM-IM block can be written as

$$\mathbf{r} = \frac{1}{\sqrt{K_q}}\mathcal{D}(e^{j\boldsymbol{\theta}})\mathbf{R}_{\mathrm{cp}}\mathbf{H}_t\mathbf{T}_{\mathrm{cp}}\mathbf{F}^H\mathbf{x} + \mathbf{w} = \frac{1}{\sqrt{K_q}}\mathcal{D}(e^{j\boldsymbol{\theta}})\mathbf{F}^H\mathbf{H}\mathbf{x} + \mathbf{w}, \tag{2.1}$$

where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_{N_q}]^T$ is the discrete PHN sample vector, $\mathbf{T}_{\mathrm{cp}} = \begin{bmatrix} \mathbf{0}_{L_{\mathrm{cp}} \times (N_q - L_{\mathrm{cp}})} & \mathbf{I}_{L_{\mathrm{cp}}} \\ & \mathbf{I}_{N_q} \end{bmatrix} \in \mathbb{R}^{(N_q + L_{\mathrm{cp}}) \times N_q}$ inserts the L_{cp}-length CP, $\mathbf{H}_t$ denotes the time-domain channel convolutional matrix consisting of the channel vector $\mathbf{h} = [h_1, \ldots, h_{L_{\mathrm{h}}}]^T \in \mathbb{C}^{L_{\mathrm{h}} \times 1}$, $\mathbf{R}_{\mathrm{cp}} = \begin{bmatrix} \mathbf{0}_{N_q \times L_{\mathrm{cp}}} & \mathbf{I}_{N_q} \end{bmatrix} \in \mathbb{R}^{N_q \times (N_q + L_{\mathrm{cp}})}$ removes the CP, and $\mathbf{w} \in \mathbb{C}^{N_q \times 1}$ is the zero-mean complex circularly symmetric Gaussian noise vector with a variance of $2\sigma^2$. In (2.1), $\mathbf{R}_{\mathrm{cp}}\mathbf{H}_t\mathbf{T}_{\mathrm{cp}} = \frac{1}{N_q}\mathbf{F}^H\mathbf{H}\mathbf{F}$ is a circular matrix and $\mathbf{H} = \mathcal{D}(\mathbf{F}\mathbf{I}_{L_{\mathrm{h}}}^0\mathbf{h})$ with $\mathbf{I}_{L_{\mathrm{h}}}^0$ being formed by the first L_{h} columns from $\mathbf{I}_{N_q}$.

As illustrated in Fig. 2.1, at the receiver, the proposed joint PHN estimation and OFDM-IM decoding algorithm is performed by: 1) estimating the PHN using the received signal sequence and the OFDM-IM symbols' estimates from the previous iteration; 2) updating the OFDM-IM symbols' estimates using the received signal sequence and the PHN samples' most recent estimates; 3) calculating the *a posteriori* probability density functions (pdfs) of information bits according to the IM and PSK/QAM modulation constraint; 4) extracting extrinsic information from the *a posteriori* information and then forwarding them to the LDPC SISO decoder to perform Bahl, Cocke, Jelinek, and Raviv (BCJR) decoding [73]; and 5) calculating the *a priori* pdfs of information bits using the outputs of the decoder and then constructing the sparse *a priori* pdfs of the OFDM-IM symbols, which are further exploited for updating the PHN samples' and OFDM-IM symbols' estimates for the next iteration.

2.3 The Proposed Joint PHN Estimation and OFDM-IM Symbol Detection Algorithms

2.3.1 Probabilistic Representation and Factor Graph

We employ a Wiener process [37–39], which is widely used for modelling a free-running oscillator, in order to approximate the variation of PHN, i.e.,

$$\theta_n = \theta_{n-1} + \triangle_n, \quad n = 2, \ldots, N_q, \tag{2.2}$$

where $\triangle_n$ is a zero-mean white Gaussian noise process with a variance of $\sigma_{\triangle}^2$. Based on (2.2) and on the received signal model in (2.1), we obtain the joint *a posteriori* pdf of $\mathbf{x}$ and $\boldsymbol{\theta}$

$$p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{r}) \propto f_C(\mathbf{c}, \mathbf{b}) \prod_{g=1}^{G} f_{M,g}(\mathbf{x}_g, \mathbf{c}_g) P(\mathbf{x}) p(\theta_1) \prod_{n=2}^{N_q} f_{\theta_n}(\theta_n, \theta_{n-1}) \cdot \prod_{n=1}^{N_q} f_{r_n}(z_n, \theta_n) f_{\delta_n}(z_n, \mathbf{x}), \tag{2.3}$$

where $f_C(\mathbf{c}, \mathbf{b})$ represents the coding constraint; $f_{M,g}$ denotes the IM and PSK/QAM modulation constraint; $P(\mathbf{x})$ is the *a priori* pdf of the data symbol $\mathbf{x}$; $p(\theta_1)$ is the *a*

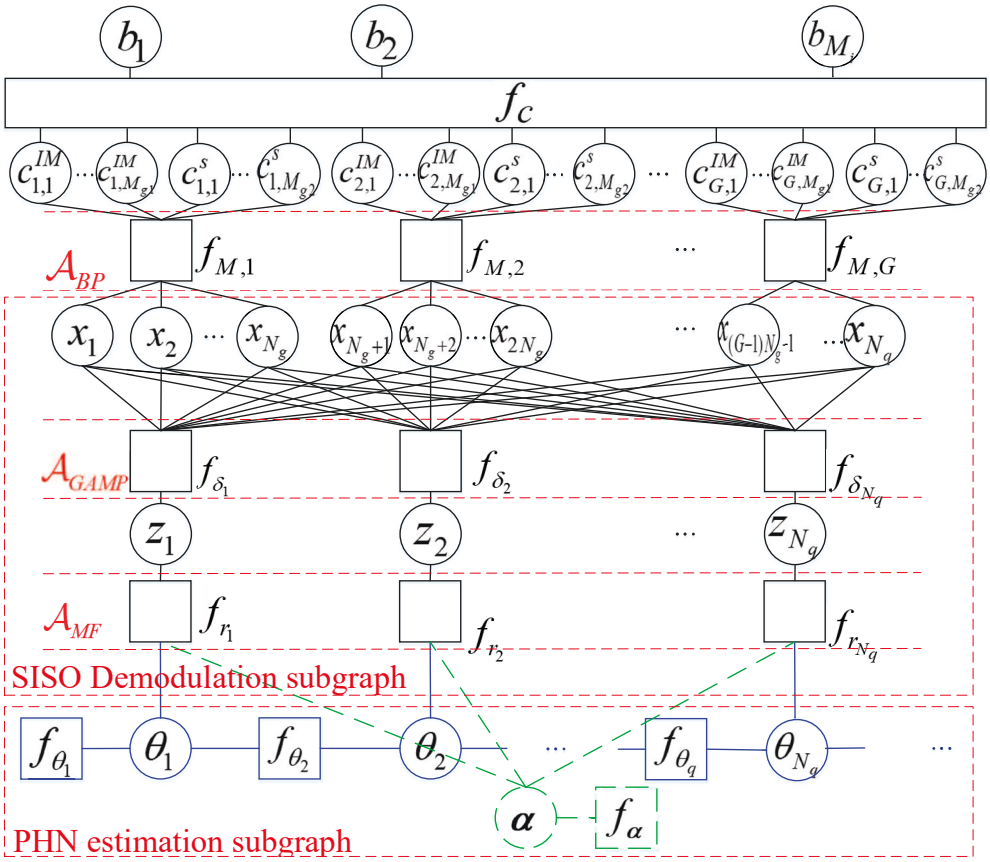

Figure 2.2: Factor graph representation of the probabilistic model (2.3), where the solid lines and dashed lines in the PHN estimation subgraph are associated with the Wiener model and DCT model, respectively.

priori pdf of θ_1 and $f_{\theta_n}(\theta_n, \theta_{n-1}) \propto \mathcal{N}(\theta_n; \theta_{n-1}, \sigma_{\triangle}^2)$. In (2.3), the additional variable z_n is defined as $z_n \triangleq \mathbf{\Lambda}_{n,:}\mathbf{x}$, where

$$\mathbf{\Lambda} \triangleq \frac{1}{\sqrt{K_q}}\mathbf{F}^H\mathbf{H}. \tag{2.4}$$

Accordingly, the relationship between the variable z_n and the vector $\mathbf{x}$ can be written as $f_{\delta_n}(z_n, \mathbf{x}) \triangleq p(z_n|\mathbf{x}) = \delta(z_n - \mathbf{\Lambda}_{n,:}\mathbf{x})$, and the likelihood function can be calculated by $f_{r_n}(z_n, \theta_n) \triangleq p(r_n|z_n, \theta_n) = \mathcal{N}_c(r_n; z_n e^{j\theta_n}, 2\sigma^2)$. The factor graph corresponding to (2.3) is shown in Fig. 2.2, where the solid lines in the PHN estimation subgraph are associated with the Wiener model.

We classify the factor nodes into three disjoint subsets, i.e., $\mathcal{A}_{\mathbf{BP}} \triangleq \{\{f_{M,g}\}_{g=1}^{G}, \{f_{\theta_n}\}_{n=1}^{N_q}\}$, $\mathcal{A}_{\mathbf{MF}} \triangleq \{f_{r_n}\}_{n=1}^{N_q}$ and $\mathcal{A}_{\mathbf{GAMP}} \triangleq \{f_{\delta_n}\}_{n=1}^{N_q}$. The messages can be readily calculated by splitting the factor graph in this way, because BP is compatible with hard constraints [67], whilst MF has simple message passing update rules for

continuous variables belonging to an exponential family [74], and GAMP works well for estimating a vector transformed by a known linear mixing matrix [75, 76].

2.3.2 Soft Demodulation and Decoding for OFDM-IM Symbols

In frequency-selective fading channels which are free from PHN, the system model is formulated in the frequency domain, where frequency-domain received signals are subjected to OFDM-IM detection. In this case, the complexity of the optimal receiver is on the order of $\mathcal{O}(G2^{M_g})$ [13], since the ML detector can be applied to each subblock independently. In the presence of correlated PHN, however, the received IM symbols become correlated and ICI occurs. In this case, the optimal detector has to consider all the legitimate IM-symbol combinations, hence the complexity order becomes $\mathcal{O}(2^{GM_g})$, which grows exponentially with G, K_g, N_g and S. Similarly to classic OFDM [37–39], the received time-domain signals are directly used for jointly estimating the PHN and for detecting the OFDM-IM symbols. As seen in (2.1), the frequency-domain OFDM-IM symbols are coupled by the transform matrix $\mathcal{D}(e^{j\boldsymbol{\theta}})\mathbf{F}^H\mathbf{H}$. The relationship between the input and output is both nonlinear and unknown due to the existence of PHN, which imposes challenges on the detection. Moreover, the complexity of optimal OFDM-IM detection becomes much higher than that of the classic OFDM symbols in both cases, since the receiver has to detect both the subcarrier index-modulated and the PSK/QAM modulated symbols. In the following, we conceive low-complexity methods for joint PHN estimation and decoding of the OFDM-IM symbols affected by PHN.

A straightforward technique of reducing the complexity is to treat x_n, $n = 1, \ldots, N_q$, as Gaussian variables [37, 38, 77, 78]. However, significant performance loss is imposed by this method, because x_n is actually a discrete variable and $\mathbf{x}_g$ has a sparse structure. Hence we invoke the GAMP algorithm [66] for decoupling the signal vector estimation problem into N_q scalar problems, whilst taking into account its discrete and sparse nature. Unfortunately, the existing GAMP method of [66] cannot be directly invoked for OFDM-IM detection, because the nonlinear

transform matrix imposing the PHN has to be known, which is not the case. Secondly, the sparsity of the signal cannot be readily exploited by the original GAMP [66]. Thirdly, it is more difficult to take full account of the activation pattern constraint if more than one subcarrier of the gth subgroup are activated. To circumvent these problems, we invoke the MF method to approximate the likelihood $p(r_n|z_n)$ for exploiting the available *a priori* information of the frequency-domain symbols provided by the LDPC SISO decoder. Hence appropriately adopt the GAMP algorithm for updating the *a posteriori* pdfs and the extrinsic information of frequency-domain symbols using our turbo receiver.

Specifically, instead of using the independent *a priori* information of the IM-symbols in the form $P(\mathbf{x}) = \prod_{n=1}^{N_q} P(x_n)$, we take into account their correlation for both the IM symbols and PSK/QAM symbols in the gth IM-symbol, using the extrinsic information gleaned from the LDPC SISO decoder (refer to [70] for details about the turbo detection principle), which can be written as $\{P(c_{g,1}^{\mathrm{IM}}), \ldots, P(c_{g,M_{g1}}^{\mathrm{IM}}), P(c_{g,1}^{\mathcal{S}}), \ldots, P(c_{g,M_{g2}}^{\mathcal{S}})\}$.

According to the associated IM and PSK/QAM modulation constraint, we have

$$\begin{aligned} P\left(\mathbf{x}_g^{\mathrm{AP}} = \boldsymbol{\omega}_\gamma\right) &= \prod_{i=1}^{M_{g1}} P\left(c_{g,i}^{\mathrm{IM}} = c_i^{\boldsymbol{\omega}_\gamma}\right), \quad \gamma = 1, \ldots, \Upsilon, \\ P(x_{g,d}(k) = s_m) &= \prod_{i=1}^{\log_2 S} P\left(c_{g,(k-1)\log_2 S+i}^{\mathcal{S}} = c_i^{s_m}\right), \\ & \qquad m = 1, \ldots, S, k = 1, \ldots, K_g. \end{aligned} \tag{2.5}$$

Thus, the *a priori* pdf of $\mathbf{x}$ provided by the LDPC SISO decoder can be formulated as

$$\begin{aligned} P(\mathbf{x}) &= \prod_{g=1}^{G} P_{\mathbf{x}_g}(\mathbf{x}_g), \\ P_{\mathbf{x}_g}(\mathbf{x}_g) &= \sum_{\gamma=1}^{\Upsilon} P\left(\mathbf{x}_g^{\mathrm{AP}} = \boldsymbol{\omega}_\gamma\right) \left[\prod_{i\in\mathcal{I}^\gamma}\left(\sum_{s\in\mathcal{S}} \delta(x_i - s) P(x_i)\right) \prod_{\jmath\in\bar{\mathcal{I}}^\gamma} \delta(x_\jmath)\right], \end{aligned} \tag{2.6}$$

where $i, \jmath \in I_g \triangleq \{(g-1)N_g + 1 : gN_g\}$, $\mathcal{I}^\gamma$ is the support set of vector $\boldsymbol{\omega}_\gamma$, i.e., $\mathcal{I}^\gamma = \{i|\omega_{\gamma,i} = 1, i \in I_g\}$, $\bar{\mathcal{I}}^\gamma$ is the complementary set of $\mathcal{I}^\gamma$, and $x_i \in \{x_{g,d}(1), \ldots, x_{g,d}(K_g)\}$.

$$b(x_i = s_m) = \sum_{\gamma: i \in \mathcal{I}^\gamma} P\left(\mathbf{x}_g^{\mathrm{IM}} = \boldsymbol{\omega}_\gamma\right) \exp\left\{-\frac{|s_m - \hat{\vartheta}_i|^2}{v_{\vartheta_i}}\right\} P(x_i = s_m) \cdot \prod_{\jmath \in \mathcal{I}^\gamma, \jmath \neq i} \left(\sum_{s \in \mathcal{S}} \exp\left\{-\frac{|s - \hat{\vartheta}_\jmath|^2}{v_{\vartheta_\jmath}}\right\} P(x_\jmath = s)\right) \prod_{k \in \bar{\mathcal{I}}^\gamma} \exp\left\{-\frac{|\hat{\vartheta}_k|^2}{v_{\vartheta_k}}\right\}. \tag{2.7}$$

$$b(x_i = 0) = \exp\left\{-\frac{|\hat{\vartheta}_i|^2}{v_{\vartheta_i}}\right\} \cdot \sum_{\gamma: i \in \bar{\mathcal{I}}^\gamma} P\left(\mathbf{x}_g^{\mathrm{IM}} = \boldsymbol{\omega}_\gamma\right) \prod_{\jmath \in \bar{\mathcal{I}}^\gamma, \jmath \neq i} \exp\left\{-\frac{|\hat{\vartheta}_\jmath|^2}{v_{\vartheta_\jmath}}\right\} \cdot \prod_{k \in \mathcal{I}^\gamma} \left(\sum_{s \in \mathcal{S}} \exp\left\{-\frac{|s - \hat{\vartheta}_k|^2}{v_{\vartheta_k}}\right\} P(x_k = s)\right). \tag{2.8}$$

Given (2.6) and the message $m_{x_i \to f_{M,g}}(x_i) \propto \mathcal{N}_c(x_i; \hat{\tau}_i, v_{\tau_i})$, which will be later updated according to (2.21) and (2.22), we can calculate the *a posteriori* pdf $b(x_i)$ of x_i by (2.7) for $x_i = s_m, s_m \in \mathcal{S}$, and by (2.8) for $x_i = 0$.

After further algebra in manipulations, we arrive at

$$b(x_i = s_m) = \frac{D(x_i = s_m)}{\sum_{s \in \mathcal{S}} D(x_i = s) + E}, \tag{2.9}$$

for $x_i = s_m, s_m \in \mathcal{S}$, and

$$b(x_i = 0) = \frac{E}{\sum_{s \in \mathcal{S}} D(x_i = s) + E}, \tag{2.10}$$

for $x_i = 0$, where $D(x_i)$ and E are given by (2.11) and (2.12), respectively. The *a posteriori* mean and variance of x_i are formulated as

$$\hat{x}_i = \sum_{s \in \mathcal{S}} s b(x_i = s), \quad v_{x_i} = \sum_{s \in \mathcal{S}} |s|^2 b(x_i = s) - |\hat{x}_i|^2. \tag{2.13}$$

Note that there is no *a priori* information arriving from the decoder during the first

$$D(x_i = s_m) = \sum_{\gamma: i \in \mathcal{I}^\gamma} P\left(\mathbf{x}_g^{\mathrm{IM}} = \boldsymbol{\omega}_\gamma\right) \exp\left\{-\frac{|s_m|^2 - 2\Re\{s_m \hat{\vartheta}_i^*\}}{v_{\vartheta_i}}\right\} P(x_i = s_m) \cdot \prod_{\jmath \in \mathcal{I}^\gamma, \jmath \neq i} \left(\sum_{s \in \mathcal{S}} \exp\left\{-\frac{|s|^2 - 2\Re\{s \hat{\vartheta}_\jmath^*\}}{v_{\vartheta_\jmath}}\right\} P(x_\jmath = s)\right). \tag{2.11}$$

$$E = \sum_{\gamma: i \in \bar{\mathcal{I}}^{\gamma}} P\left(\mathbf{x}_g^{\mathrm{IM}} = \boldsymbol{\omega}_{\gamma}\right) \prod_{\jmath \in \mathcal{I}^{\gamma}} \left(\sum_{s \in \mathcal{S}} \exp\left\{ -\frac{|s|^2 - 2\Re\{s\hat{\vartheta}_{\jmath}^{*}\}}{v_{\vartheta_{\jmath}}} \right\} P(x_{\jmath} = s) \right). \tag{2.12}$$

iteration, and the *a priori* pdf of $\mathbf{x}_g$ can be directly written as

$$P_{\mathbf{x}_g}^{0}(\mathbf{x}_g) = \frac{1}{\Upsilon} \sum_{\gamma=1}^{\Upsilon} \left[\prod_{i \in \mathcal{I}^{\gamma}} \left(\frac{1}{M} \sum_{s \in \mathcal{S}} \delta(x_i - s) \right) \prod_{\jmath \in \bar{\mathcal{I}}^{\gamma}} \delta(x_{\jmath}) \right].$$

Accordingly, the *a posteriori* mean and variance of x_i for the first iteration is given by

$$\hat{x}_i^0 = \sum_{s \in \mathcal{S}} s \frac{||\boldsymbol{\Omega}_{i,1}||_0}{||\boldsymbol{\Omega}||_0 M}, \quad v_{x_i}^0 = \frac{1}{M} \sum_{s \in \mathcal{S}} |s|^2 \frac{||\boldsymbol{\Omega}_{i,1}||_0}{||\boldsymbol{\Omega}||_0} - |\hat{x}_i^0|^2, \tag{2.14}$$

where $\boldsymbol{\Omega}_{i,1} = \{\boldsymbol{\omega}_{\gamma} | \omega_{\gamma,i} = 1, \gamma = 1, \ldots, \Upsilon\}$.

By using the GAMP algorithm ([21, Eqs. (2a) and (2b)]), the message $m_{f_{\delta_n} \to z_n}(z_n)$ $\propto \mathcal{N}_c(z_n; \hat{\xi}_n, v_{\xi_n})$ from the factor node f_{δ_n} to the variable node z_n is updated by

$$v_{\xi_n} = \sum_i |\Lambda_{n,i}|^2 v_{x_i}, \tag{2.15}$$

$$\hat{\xi}_n = \sum_i \Lambda_{n,i} \hat{x}_i - \hat{\zeta}_n v_{\xi_n}, \tag{2.16}$$

where $\hat{\zeta}_n$ related to the *a posteriori* mean of z_n will be later updated according to (2.20). In the original GAMP, the function F_{out} defined in [66] is given by $F_{\mathrm{out}}(z_n, \hat{\xi}_n, v_{\xi_n}, r_n) = f_{\mathrm{out}} - \frac{1}{v_{\xi_n}} |z_n - \hat{\xi}_n|^2$, where $f_{\mathrm{out}} = \log p(r_n | z_n)$ [21, Eq. (7c)]. However, the relationship between z_n and r_n is nonlinear and unknown due to the presence of PHN. To this end, we use the MF method [65] for calculating the message $m_{f_{r_n} \to z_n}(z_n)$ forwarded from the factor node f_{r_n} to the variable node z_n, and use it for approximating $p(r_n | z_n)$. Given the *a posteriori* pdf of θ_n, i.e., $b(\theta_n) \propto \mathcal{N}(\theta_n; \hat{\theta}_n, v_{\theta_n})$, which will be later updated in (2.35), the message $m_{f_{r_n} \to z_n}(z_n)$ is calculated following the MF rule [20, Eq. (18)] and then approximated with the aid of its second-order Taylor expansion as:

$$\begin{aligned} m_{f_{r_n} \to z_n}(z_n) &\propto \exp\left\{ E_{b(\theta_n)} \left[\log f_{r_n}(z_n, e^{j\theta_n}) \right] \right\} \\ &\approx \mathcal{N}_c(z_n; \hat{\eta}_n, v_{\eta_n}), \end{aligned} \tag{2.17}$$

where

$$\hat{\eta}_n = r_n e^{-j\hat{\theta}_n}(1 - \frac{1}{2}v_{\theta_n}), \quad v_{\eta_n} = 2\sigma^2. \tag{2.18}$$

Bearing in mind that F_{out} is equivalent to the belief of z_n in its log-domain based on the probabilistic understanding of message passing, the *a posteriori* pdf of z_n can be expressed as $b(z_n) \propto \mathcal{N}_c(z_n; \hat{z}_n, v_{z_n})$, where

$$v_{z_n} = (v_{\xi_n}^{-1} + v_{\eta_n}^{-1})^{-1}, \quad \hat{z}_n = v_{z_n}(\hat{\xi}_n v_{\xi_n}^{-1} + \hat{\eta}_n v_{\eta_n}^{-1}). \tag{2.19}$$

The output scalar function $g_{\text{out}}(\hat{\xi}_n, v_{\xi_n}, \hat{\eta}_n, v_{\eta_n})$ is updated by ([21, Eqs. (3a) and (3b)])

$$\begin{aligned} \hat{\zeta}_n &= g_{\text{out}}(\hat{\xi}_n, v_{\xi_n}, \hat{\eta}_n, v_{\eta_n}) = \frac{\hat{z}_n - \hat{\xi}_n}{v_{\xi_n}} = \frac{\hat{\eta}_n - \hat{\xi}_n}{v_{\eta_n} + v_{\xi_n}}, \\ v_{\zeta_n} &= -\frac{\partial}{\partial \hat{\xi}_n} g_{\text{out}}(\hat{\xi}_n, v_{\xi_n}, \hat{\eta}_n, v_{\eta_n}) = \frac{1}{v_{\xi_n}}\left(1 - \frac{v_{z_n}}{v_{\xi_n}}\right). \end{aligned} \tag{2.20}$$

In the GAMP algorithm ([21, Eqs. (4a) and (4b)]), the variance and the mean of the message $m_{x_i \to f_{M,g}}(x_i) \propto \mathcal{N}_c(x_i; \hat{\vartheta}_i, v_{\vartheta_i})$ are calculated by

$$v_{\tau_i} = \left(\sum_n |\Lambda_{n,i}|^2 v_{\zeta_n}\right)^{-1}, \tag{2.21}$$

$$\hat{\tau}_i = v_{\tau_i} \sum_n \hat{\zeta}_n \Lambda_{n,i}^* + \hat{x}_i. \tag{2.22}$$

Note that the complexity of the structured GAMP algorithm is $\mathcal{O}(N_q^2)$, which is dominated by the matrix-vector multiplications in (2.15), (2.16), (2.21), and (2.22). This complexity can be reduced to $\mathcal{O}(N_q \log N_q)$ thanks to the FFT. Specifically, we can rewrite (2.15) as

$$\mathbf{v}_\xi = \frac{1}{K_q}|\mathbf{F}^H|^2 \mathbf{H}^H \mathbf{H} \mathbf{v}_x \approx \frac{1}{K_q} \sum_i |h_i|^2 v_{x_i} \mathbf{1}_{N_q \times 1}, \tag{2.23}$$

where $\mathbf{v}_\xi = [v_{\xi_1}, \ldots, v_{\xi_{N_q}}]^T$ and $\mathbf{v}_x = [v_{x_1}, \ldots, v_{x_{N_q}}]^T$. Furthermore, we have

$$\hat{\boldsymbol{\xi}} = \frac{1}{\sqrt{K_q}} \mathbf{F}^H \mathbf{H} \hat{\mathbf{x}} - \hat{\boldsymbol{\zeta}} \odot \mathbf{v}_\xi, \tag{2.24}$$

where $\hat{\boldsymbol{\xi}} = [\hat{\xi}_1, \ldots, \hat{\xi}_{N_q}]^T$, $\hat{\mathbf{x}} = [\hat{x}_1, \ldots, \hat{x}_{N_q}]^T$, and $\hat{\boldsymbol{\zeta}} = [\hat{\zeta}_1, \ldots, \hat{\zeta}_{N_q}]^T$. Similarly, (2.21) can also be represented in a vectorial form, i.e.,

$$\mathbf{v}_\tau = \mathbf{1}_{N_q \times 1} ./ (\frac{1}{K_q} \mathbf{H}^H \mathbf{H} |\mathbf{F}|^2 \mathbf{v}_\zeta) \approx \frac{K_q}{\bar{v}_\zeta} (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{1}_{N_q \times 1}, \tag{2.25}$$

where $\mathbf{v}_\tau = [v_{\tau_1}, \ldots, v_{\tau_{N_q}}]^T$, $\mathbf{v}_\zeta = [v_{\zeta_1}, \ldots, v_{\zeta_{N_q}}]^T$, and $\bar{v}_\zeta = \sum_n v_{\zeta_n}$. Hence, (2.22) is reduced to the following equation

$$\hat{\boldsymbol{\tau}} = \frac{1}{\sqrt{K_q}} \mathbf{H}^H \mathbf{F} \hat{\boldsymbol{\zeta}} \odot \mathbf{v}_\tau + \hat{\mathbf{x}}, \tag{2.26}$$

where $\hat{\boldsymbol{\tau}} = [\hat{\tau}_1, \ldots, \hat{\tau}_{N_q}]^T$ and $\hat{\boldsymbol{\zeta}} = [\hat{\zeta}_1, \ldots, \hat{\zeta}_{N_q}]^T$. Fortunately, some of the matrix-vector calculations within the GAMP iterations can be efficiently performed by the FFT. The complexity of the specific matrix-vector calculations relying on diagonal matrices is low.

Given the latest updated message $m_{x_i \to f_{M,g}}(x_i)$ in (2.21) and (2.22), the belief of x_i can be updated by (2.9) and (2.10). Meanwhile, we can obtain the belief of $\mathbf{x}_g$ as

$$b(\mathbf{x}_g) \propto P_{\mathbf{x}_g}(\mathbf{x}_g) \prod_{i \in I_g} \exp\left\{ -\frac{|x_i - \hat{\tau}_i|^2}{v_{\tau_i}} \right\}, \tag{2.27}$$

where $i \in I_g \triangleq \{(g-1)N_g + 1 : gN_g\}$. The extrinsic log-likelihood ratios (LLRs) of the ith subcarrier selection bit and of the jth symbol mapping bit in the gth IM-symbol, i.e. $L^{\text{extr}}(c_{g,i}^{\text{IM}})$ and $L^{\text{extr}}(c_{g,j}^{\mathcal{S}})$, respectively, can be calculated according to the turbo principle [68–70].

2.3.3 PHN Estimation in the OFDM-IM Scheme

PHN Estimation Based on the Wiener Model

The message emerging from the observation node to the PHN estimation subgraph can then be calculated by

$$m_{f_{r_n} \to \theta_n}(\theta_n) \propto \exp\left\{ E_{b(z_n)} \left[\log f_{r_n}(z_n, \theta_n) \right] \right\} \propto \left\{ \Re \left[\frac{r_n^* \hat{z}_n}{\sigma^2} e^{j\theta_n} \right] \right\}. \tag{2.28}$$

It is readily seen that the message $m_{f_{r_n} \to \theta_n}(\theta_n)$ in (2.28) is not Gaussian due to the nonlinear effect imposed by PHN. However, this message is expected to be Gaussian, hence the messages updated in the PHN estimation subgraph are all Gaussian and thus low complexity can be achieved. We use the second-order Taylor expansion for approximating $m_{f_{r_n} \to \theta_n}(\theta_n)$, instead of directly linearizing the system model [79–81]. Given the *a posteriori* mean $\hat{\theta}_n$ and variance v_{θ_n} obtained from the previous iteration, $e^{j\theta_n}$ can be approximated as

$$e^{j\theta_n} = e^{j\hat{\theta}_n}(1 - j\hat{\theta}_n - \frac{1}{2}\hat{\theta}_n^2) - \frac{1}{2} e^{j\hat{\theta}_n} \theta_n^2 + (j + \hat{\theta}_n) e^{j\hat{\theta}_n} \theta_n. \tag{2.29}$$

Thus, $m_{f_{r_n\to\theta_n}}(\theta_n)$ can be approximated by $\mathcal{N}(\theta_n;\hat{\theta}_n^{\downarrow},v_{\theta_n^{\downarrow}})$, with

$$\begin{aligned} v_{\theta_n^{\downarrow}} &= 1/\Re\left[\frac{r_n^*\hat{z}_n}{\sigma^2}e^{j\hat{\theta}_n}\right], \\ \hat{\theta}_n^{\downarrow} &= v_{\theta_n^{\downarrow}}\Re\left[\frac{r_n^*\hat{z}_n}{\sigma^2}(j+\hat{\theta}_n)e^{j\hat{\theta}_n}\right]. \end{aligned} \tag{2.30}$$

Given this Gaussian message forwarded from the MF subgraph to the PHN estimation subgraph, both $m_{f_{\theta_{n-1}\to\theta_n}}(\theta_n)$ and $m_{f_{\theta_{n+1}\to\theta_n}}(\theta_n)$ are Gaussian and can be recursively updated based on the BP algorithm ([19, Eqs. (5) and (6)]) as follows:

$$\begin{aligned} m_{f_{\theta_n\to\theta_n}}(\theta_n) &\propto \int m_{f_{\theta_{n-1}\to\theta_{n-1}}}(\theta_{n-1})f_{\theta_n}(\theta_n,\theta_{n-1})m_{f_{r_{n-1}\to\theta_{n-1}}}(\theta_{n-1})\mathrm{d}\theta_{\mathrm{n-1}} \\ &\propto \mathcal{N}(\theta_n;\hat{\theta}_n^{\rightarrow},v_{\theta_n^{\rightarrow}}), \end{aligned} \tag{2.31}$$

where

$$\begin{aligned} \hat{\theta}_n^{\rightarrow} &= \frac{(v_{\theta_{n-1}^{\rightarrow}})^{-1}\hat{\theta}_{n-1}^{\rightarrow}+(v_{\theta_{n-1}^{\downarrow}})^{-1}\hat{\theta}_{n-1}^{\downarrow}}{(v_{\theta_{n-1}^{\rightarrow}})^{-1}+(v_{\theta_{n-1}^{\downarrow}})^{-1}}, \\ v_{\theta_n^{\rightarrow}} &= \sigma_{\triangle}^2+\left((v_{\theta_{n-1}^{\rightarrow}})^{-1}+(v_{\theta_{n-1}^{\downarrow}})^{-1}\right)^{-1}, \end{aligned} \tag{2.32}$$

and

$$\begin{aligned} m_{f_{\theta_{n+1}\to\theta_n}}(\theta_n) &\propto \int m_{f_{\theta_{n+2}\to\theta_{n+1}}}(\theta_{n+1})f_{\theta_{n+1}}(\theta_{n+1},\theta_n)m_{f_{r_{n+1}\to\theta_{n+1}}}(\theta_{n+1})\mathrm{d}\theta_{\mathrm{n+1}} \\ &\propto \mathcal{N}(\theta_n;\hat{\theta}_n^{\leftarrow},v_{\theta_n^{\leftarrow}}), \end{aligned} \tag{2.33}$$

where

$$\begin{aligned} \hat{\theta}_n^{\leftarrow} &= \frac{(v_{\theta_{n+1}^{\leftarrow}})^{-1}\hat{\theta}_{n+1}^{\leftarrow}+(v_{\theta_{n+1}^{\downarrow}})^{-1}\hat{\theta}_{n+1}^{\downarrow}}{(v_{\theta_{n+1}^{\leftarrow}})^{-1}+(v_{\theta_{n+1}^{\downarrow}})^{-1}}, \\ v_{\theta_n^{\leftarrow}} &= \sigma_{\triangle}^2+\left((v_{\theta_{n+1}^{\leftarrow}})^{-1}+(v_{\theta_{n+1}^{\downarrow}})^{-1}\right)^{-1}. \end{aligned} \tag{2.34}$$

Given (2.28), (2.31) and (2.33), the belief of θ_n is expressed as ([19, Eq. (3)]):

$$\begin{aligned} b(\theta_n) &\propto m_{f_{r_n\to\theta_n}}(\theta_n)m_{f_{\theta_n}\to\theta_n}(\theta_n)m_{f_{\theta_{n+1}}\to\theta_n}(\theta_n) \\ &\propto \mathcal{N}(\theta_n;\hat{\theta}_n,v_{\theta_n}), \end{aligned} \tag{2.35}$$

where

$$\begin{aligned} v_{\theta_n} &= (v_{\theta_n^{\rightarrow}}^{-1}+v_{\theta_n^{\downarrow}}^{-1}+v_{\theta_n^{\uparrow}}^{-1})^{-1}, \\ \hat{\theta}_n &= v_{\theta_n}(v_{\theta_n^{\rightarrow}}^{-1}\hat{\theta}_n^{\rightarrow}+v_{\theta_n^{\downarrow}}^{-1}\hat{\theta}_n^{\downarrow}+v_{\theta_n^{\uparrow}}^{-1}\hat{\theta}_n^{\uparrow}). \end{aligned} \tag{2.36}$$

The message passing schedule for our proposed joint PHN estimation and decoding algorithm based on the GAMP method relying on the Wiener model of the OFDM-IM scheme, which is referred to as GAMP-Wiener, is summarized in Algorithm 1.

Algorithm 1 The proposed joint PHN estimation and decoding algorithm based on the GAMP method relying on the Wiener model.

1: Initialization: $t = 0$ (iteration index)

2: For $n = 1, \ldots, N_q$, compute $\hat{x}_n^0$ and $v_{x_n}^0$; set $\hat{\zeta}_n^0 = 0$, $\hat{\eta}_n^0 = r_n$ and $v_{\eta_n}^0 = 2\sigma^2$.

3: Set $t = 1$ and repeat the following steps until $\sum_n |\hat{x}_n^{t+1} - \hat{x}_n^t|^2 < \epsilon \sum_n |\hat{x}_n^t|^2$ or $t > T_{\max}$, where ϵ is a pre-specified error tolerance and $T_{\max}$ is the maximum number of iterations.

4: − PHN estimation step: For $n = 1, \ldots, N_q$, compute $\{\mathbf{v}_\xi^t, \hat{\boldsymbol{\xi}}^t\}$ and $\{\hat{z}_n^t, v_{z_n}^t\}$ from (2.23), (2.24) and (2.19), respectively.

5: For $n = 1, \ldots, N_q$, compute $\{\hat{\theta}_n^{\downarrow}, v_{\theta_n^{\downarrow}}\}$, $\{\hat{\theta}_n^{\rightarrow}, v_{\theta_n^{\rightarrow}}\}$, $\{\hat{\theta}_n^{\leftarrow}, v_{\theta_n^{\leftarrow}}\}$ and $\{\hat{\theta}_n^t, v_{\theta_n}^t\}$ from (2.30), (2.32), (2.34) and (2.36).

6: − Decoupling step: For $n = 1, \ldots, N_q$, compute $\{\hat{\eta}_n^t, v_{\eta_n}^t\}$, $\{\hat{z}_n^t, v_{z_n}^t\}$, $\{\hat{\zeta}_n^t, v_{\zeta_n}^t\}$, $\{\mathbf{v}_\tau^t, \hat{\boldsymbol{\tau}}^t\}$ from (2.18), (2.19), (2.20), (2.25) and (2.26), respectively.

7: − Decoding step: For $g = 1 \ldots, G$, $i = 1, \ldots, M_{g1}$, $\jmath = 1, \ldots, M_{g2}$, $L^{\text{extr}}(c_{g,i}^{\text{IM}})$ and $L^{\text{extr}}(c_{g,\jmath}^{\mathcal{S}}) \rightarrow$ Decoder.

8: Compute (2.5) by using the information from the decoder.

9: − Denoising step: For $n = 1, \ldots, N_q$, compute $\{\hat{x}_n^t, v_{x_n}^t\}$ from (2.13).

PHN Estimation Based on the DCT Model

Inspired by the idea of reducing the number of unknown PHN samples and improving the degree of parallelism, we model the PHN variation as a truncated DCT expansion [71], i.e. the vector $\mathbf{u} \triangleq e^{\jmath\boldsymbol{\theta}}$ is approximately represented by

$$\mathbf{u} = \boldsymbol{\Phi}\boldsymbol{\alpha}, \tag{2.37}$$

where $\mathbf{\Phi}$ is the DCT matrix with size $N_q \times L_{\mathrm{PN}}$, and $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_{L_{\mathrm{PN}}}]^T$ collects the DCT coefficients to be estimated[c]. In this way, the problem of estimating the N_q dimensional PHN vector has been transformed to the estimation of L_{PN} DCT coefficients. Based on (2.1) and (2.37), the received data vector can be alternatively expressed as:

$$\mathbf{r} = \frac{1}{\sqrt{K_q}} \mathcal{D}(\mathbf{\Phi}\boldsymbol{\alpha}) \mathbf{F}^H \mathbf{H} \mathbf{x} + \mathbf{w}. \tag{2.38}$$

The PHN estimation subgraph corresponding to the DCT model is shown on top of the Wiener model factor graph representation in Fig. 2.2 using dashed lines. The message arriving from the observation node to the DCT coefficient vector $\boldsymbol{\alpha}$ is formulated as:

$$\begin{aligned} m_{f_{r_n} \to \boldsymbol{\alpha}}(\boldsymbol{\alpha}) &\propto \exp\left\{ E_{b(z_n)} [\log f_{r_n(\boldsymbol{\alpha}, z_n)}] \right\} \\ &\propto \exp\left\{ \frac{2\Re\{r_n^* \hat{z}_n (\mathbf{\Phi}\boldsymbol{\alpha})_n\} - |(\mathbf{\Phi}\boldsymbol{\alpha})_n|^2 (|\hat{z}_n|^2 + v_{z_n})}{2\sigma^2} \right\}. \end{aligned}$$

Thus, the *a posteriori* pdf of $\boldsymbol{\alpha}$ is given by

$$\begin{aligned} b(\boldsymbol{\alpha}) &\propto p(\boldsymbol{\alpha}) \prod_{n=1}^{N_q} m_{f_{r_n} \to \boldsymbol{\alpha}}(\boldsymbol{\alpha}) \\ &\propto \exp\left\{ \frac{2\Re\{(\mathbf{r} \odot \hat{\mathbf{z}}^*)^H \mathbf{\Phi}\boldsymbol{\alpha}\} - \boldsymbol{\alpha}^H \mathbf{\Phi}^H \mathcal{D}(\hat{\boldsymbol{\varrho}}_z) \mathbf{\Phi}\boldsymbol{\alpha}}{2\sigma^2} \right\}, \end{aligned}$$

where $\hat{\boldsymbol{\varrho}}_z = [|\hat{z}_1|^2 + v_{z_1}, \ldots, |\hat{z}_{N_q}|^2 + v_{z_{N_q}}]^T$, while its covariance matrix and mean vector are

$$\mathbf{V}_{\boldsymbol{\alpha}} = 2\sigma^2 (\mathbf{\Phi}^H \mathcal{D}(\hat{\boldsymbol{\varrho}}_z) \mathbf{\Phi})^{-1}, \quad \hat{\boldsymbol{\alpha}} = \frac{1}{2\sigma^2} \mathbf{V}_{\boldsymbol{\alpha}} \mathbf{\Phi}^H (\mathbf{r} \odot \hat{\mathbf{z}}^*). \tag{2.39}$$

Note that the computation of (2.39) involves a nontrivial matrix inversion, which has a complexity order of $\mathcal{O}(L_{\mathrm{PN}}^3)$.

Given $b(\boldsymbol{\alpha})$, the message $m_{f_{r_n} \to z_n}(z_n)$ is expressed as:

$$\begin{aligned} m_{f_{r_n} \to z_n}(z_n) &\propto \exp\{ E_{b(\boldsymbol{\alpha})} [\log f_{r_n}(z_n, \boldsymbol{\alpha})] \} \\ &\approx \mathcal{N}_c(z_n; \hat{\eta}_n, v_{\eta_n}), \end{aligned} \tag{2.40}$$

[c]The design of the length of vector $\mathbf{u}$ is flexible. The selection of N_q here is just for the sake of simplicity.

where

$$\hat{\eta}_n = \frac{r_n}{\hat{\varrho}_{u_n}}(\mathbf{\Phi}\hat{\boldsymbol{\alpha}})_n^*, \quad v_{\eta_n} = 2\sigma^2/\hat{\varrho}_{u_n}, \tag{2.41}$$

with

$$\hat{\varrho}_{u_n} = \sum_{i=1}^{L_{\mathrm{PN}}} |\Phi_{n,i}|^2 (V_{i,i} + |\hat{\alpha}_i|^2) + \sum_{i=1}^{L_{\mathrm{PN}}} \sum_{\jmath=1,\jmath>i}^{L_{\mathrm{PN}}} \Phi_{n,i}\Phi_{n,\jmath} 2\Re(V_{i,\jmath} + \hat{\alpha}_i^*\hat{\alpha}_\jmath). \tag{2.42}$$

Actually, the complexity of both (2.39) and (2.42) is low. As it will be demonstrated by our simulation results, an attractive BER performance can be attained when L_{PN} is selected to be sufficiently small compared to N_q. To further reduce this complexity, we approximate $\mathcal{D}(\hat{\boldsymbol{\varrho}}_z)$ by $\hat{\varrho}_z \mathbf{I}_{N_q}$, where $\hat{\varrho}_z$ is the average of the diagonal elements of $\mathcal{D}(\hat{\boldsymbol{\varrho}}_z)$. In this way, $\mathbf{V}_{\boldsymbol{\alpha}}$ becomes a diagonal matrix, with $2\sigma^2/\hat{\varrho}_z$ being the diagonal element since $\mathbf{\Phi}^H\mathbf{\Phi} = \mathbf{I}_{L_{\mathrm{PN}}}$.

The message passing schedule for the proposed joint PHN estimation and decoding algorithm based on the GAMP method using the DCT model for the OFDM-IM scheme, which is referred to as GAMP-DCT, is summarized in Algorithm 2.

Algorithm 2 The proposed joint PHN estimation and decoding algorithm based on the GAMP method using the DCT model.

1: The same as GAMP-Wiener except for the following modifications imposed on the PHN estimation step and the decoupling step:
2: − PHN estimation step: For $n = 1, \ldots, N_q$, compute $\{\mathbf{v}_\xi^t, \hat{\boldsymbol{\xi}}^t\}$ and $\{\hat{z}_n^t, v_{z_n}^t\}$ from (2.23), (2.24) and (2.19), respectively.
3: Compute $\{\mathbf{V}_{\boldsymbol{\alpha}}^t, \hat{\boldsymbol{\alpha}}^t\}$ from (2.39).
4: − Decoupling step: For $n = 1, \ldots, N_q$, compute $\{\hat{\eta}_n^t, v_{\eta_n}^t\}$, $\{\hat{z}_n^t, v_{z_n}^t\}$, $\{\hat{\zeta}_n^t, v_{\zeta_n}^t\}$, $\{\mathbf{v}_\tau^t, \hat{\boldsymbol{\tau}}^t\}$ from (2.41), (2.19), (2.20), (2.25) and (2.26), respectively.

2.4 Amalgamated BP-MF Algorithm

For comparison, we also amalgamate the BP-MF algorithm of [67] and invoke it for the coded OFDM-IM system. The subgraph of the PHN estimation shown in Fig. 2.2 is also used in the BP-MF method. The difference with regard to

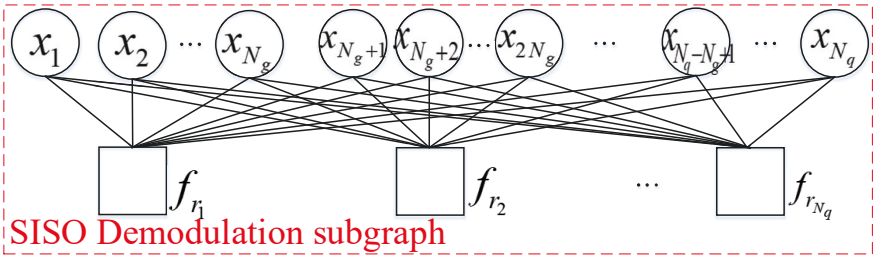

Figure 2.3: The subgraph for soft demodulation using the amalgamated BP-MF algorithm.

the proposed GAMP based algorithms is that the BP-MF technique is invoked for soft demodulation, and the corresponding SISO demodulation subgraph shown in Fig. 2.2 is replaced by that in Fig. 2.3. In contrast to the traditional OFDM system, the sparsity of the OFDM-IM symbol should be explicitly exploited by the BP-MF method, which is similar to that of the GAMP based algorithms. However, instead of using central-limit-theorem based approximations (refer to [66] for details of original GAMP) for representing the noiseless output z_n, $n = 1, \ldots, N_q$, and assuming them to be independent of each other, we use the *a posteriori* pdf of z_n to estimate the PHN effect $e^{j\theta_n}$, the BP-MF method estimates $\mathbf{x}$ directly and uses the *a posteriori* pdf of this vector for updating the PHN samples' estimates. As shown in Fig. 2.3, this part of the subgraph is densely connected. Therefore, there are multiple message-passing schedules to select. In the following, both a sequential and a parallel schedule are used. Bearing in mind that the DCT model based PHN estimation method operates in parallel, while the Wiener model based one is in a sequential recursion form, they are naturally embedded in parallel and sequential message-passing schedules, respectively. In both schedules, the extrinsic information extracted from the LDPC decoder is calculated simultaneously.

The message $m^{\mathrm{BP-MF}}_{x_i \to f_{M,g}}(x_i)$ is formulated as:

$$m^{\mathrm{BP-MF}}_{x_i \to f_{M,g}}(x_i) \propto \prod_{i=1}^{N_q} m_{f_{rn} \to x_i}(x_i) \propto \mathcal{N}_c(x_i; \underline{\hat{\tau}}_i, v_{\underline{\tau}_i}), \tag{2.43}$$

where $\underline{\hat{\tau}}_i$ and $v_{\underline{\tau}_i}$ will be later updated according to (2.47) and (2.52) in the sequential and parallel schedules, respectively. Given the extrinsic information gleaned from the SISO decoder and (2.43), we can compute the *a posteriori* pdf of x_i, $\underline{b}(x_i)$ using

(2.9) and (2.10), where (2.11) and (2.12) required in (2.9) and (2.10) are obtained upon replacing $\{\hat{\tau}_i, v_{\tau_i}\}$ by $\{\underline{\hat{\tau}}_i, v_{\underline{\tau}_i}\}$. The *a posteriori* mean and variance of x_i are obtained as

$$\underline{\hat{x}}_i = \sum_{s\in\mathcal{S}} s\underline{b}(x_i = s), \quad v_{\underline{x}_i} = \sum_{s\in\mathcal{S}} |s|^2 \underline{b}(x_i = s) - |\underline{\hat{x}}_i|^2. \tag{2.44}$$

2.4.1 Wiener Model Based Sequential BP-MF Message Passing Schedule

According to the MF rule, the message $m_{f_{r_n}\to x_i}(x_i)$ can be expressed using $\underline{b}(x_i)$ and $b(\theta_n)$ by

$$\begin{aligned} m_{f_{r_n}\to x_i}(x_i) &\propto \exp\left\{E_{b(\theta_n)\prod_{\jmath\neq i}\underline{b}(x_\jmath)}\left[\log f_{r_n}(\mathbf{x}, \theta_n)\right]\right\} \\ &\propto \mathcal{N}_c(x_i; \hat{\tau}_i^{'}, v_{\tau_i^{'}}), \end{aligned} \tag{2.45}$$

where

$$\begin{aligned} v_{\tau_i^{'}} &= 2\sigma^2/|\Lambda_{n,i}|^2, \\ \hat{\tau}_i^{'} &= \frac{v_{\tau_i^{'}}\Lambda_{n,i}^*}{2\sigma^2}\left(r_n e^{-j\hat{\theta}_n}(1 - \frac{1}{2}v_{\theta_n}) - \mathbf{\Lambda}_{n,:}\hat{\mathbf{x}}_{\bar{i}}\right). \end{aligned} \tag{2.46}$$

In (2.46), $\hat{\mathbf{x}}_{\bar{i}}$ is constructed of the *a posteriori* mean of $x_\jmath$, $\jmath = 1 \ldots N_q, \jmath \neq i$, i.e., $\hat{\mathbf{x}}_{\bar{i}} = [\hat{x}_1, \ldots, \hat{x}_{i-1}, 0, \hat{x}_{i+1}, \ldots, \hat{x}_{N_q}]^T$. Substituting (2.45) into (2.43), we have

$$\begin{aligned} v_{\underline{\tau}_i} &= 2\sigma^2\left(\mathbf{\Lambda}_{:,i}^H\mathbf{\Lambda}_{:,i}\right)^{-1}, \\ \underline{\hat{\tau}}_i &= \frac{v_{\underline{\tau}_i}}{2\sigma^2}\mathbf{\Lambda}_{:,i}^H\left(\mathcal{D}(\hat{\boldsymbol{\varrho}}_\theta)\mathbf{r} - \mathbf{\Lambda}\hat{\mathbf{x}}_{\bar{i}}\right), \end{aligned} \tag{2.47}$$

where $\hat{\boldsymbol{\varrho}}_\theta = e^{-j\hat{\boldsymbol{\theta}}} - \frac{1}{2}e^{-j\hat{\boldsymbol{\theta}}} \odot \mathbf{v}_\theta$, $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \ldots, \hat{\theta}_{N_q}]^T$ and $\mathbf{v}_\theta = [v_{\theta_1}, \ldots, v_{\theta_{N_q}}]^T$.

Using the second-order Taylor expansion of $e^{j\theta_n}$ in (2.29), the message passed on the MF subgraph to the PHN estimation subgraph can be approximated as

$$\begin{aligned} m_{f_{r_n}\to\theta_n}^{\mathrm{BP-MF}}(\theta_n) &\propto \exp\left\{E_{\underline{b}(\mathbf{x})}\left[\log f_{r_n}(\mathbf{x}, \theta_n)\right]\right\} \\ &\propto \left\{\Re\left[\frac{r_n^*\mathbf{\Lambda}_{n,:}\hat{\mathbf{x}}}{\sigma^2}e^{j\theta_n}\right]\right\} \propto \mathcal{N}(\theta_n; \hat{\theta}_n^{\downarrow}, v_{\theta_n^{\downarrow}}), \end{aligned} \tag{2.48}$$

where

$$\begin{aligned} v_{\theta_n^{\downarrow}} &= 1/\Re\left[\frac{r_n^*\mathbf{\Lambda}_{n,:}\hat{\mathbf{x}}}{\sigma^2}e^{j\hat{\theta}_n}\right], \\ \hat{\theta}_n^{\downarrow} &= v_{\theta_n^{\downarrow}}\Re\left[\frac{r_n^*\mathbf{\Lambda}_{n,:}\hat{\mathbf{x}}}{\sigma^2}(\jmath + \hat{\theta}_n)e^{j\hat{\theta}_n}\right]. \end{aligned} \tag{2.49}$$

The amalgamated BP-MF using the sequential message passing schedule for the Wiener model based joint PHN estimation and decoding algorithm, which is referred to as BP-MF-s, is summarized in Algorithm 3.

Algorithm 3 Amalgamated BP-MF algorithm for joint PHN estimation and decoding based on the Wiener model using the sequential scheduling.

1: Initialization: $t = 0$ (iteration index)
2: For $n = 1, \ldots, N_q$, compute $\hat{x}_n^0$ and $v_{x_n}^0$; set $\hat{\theta}_n = 0$.
3: Set $t = 1$ and repeat the following steps until $\sum_n |\hat{x}_n^{t+1} - \hat{x}_n^t|^2 < \epsilon \sum_n |\hat{x}_n^t|^2$ or $t > T_{\max}$.
4: $-$ PHN estimation step: For $n = 1, \ldots, N_q$, compute $\{\hat{\theta}_n^{\downarrow}, v_{\theta_n^{\downarrow}}\}$, $\{\hat{\theta}_n^{\rightarrow}, v_{\theta_n^{\rightarrow}}\}$, $\{\hat{\theta}_n^{\leftarrow}, v_{\theta_n^{\leftarrow}}\}$ and $\{\hat{\theta}_n^t, v_{\theta_n}^t\}$ from (2.49), (2.32), (2.34) and (2.36), respectively.
5: $-$ Soft demodulation step: Compute $\hat{\mathbf{w}}^t = \mathcal{D}(\hat{\boldsymbol{\varrho}}_\theta^t)\mathbf{r} - \boldsymbol{\Lambda}\hat{\mathbf{x}}^{t-1}$.
6: For $n = 1, \ldots, N_q$,
7: compute $v_{\underline{\tau}_n}^t$ from (2.47); compute $\hat{\underline{\tau}}_n^t = \frac{v_{\underline{\tau}_n}^t}{2\sigma^2}\boldsymbol{\Lambda}_{:,n}^H \hat{\mathbf{w}}^t + \hat{\underline{x}}_n^{t-1}$; compute $\{\hat{\underline{x}}_n^t, v_{\underline{x}_n}^t\}$; compute $\hat{\mathbf{w}}^t = \hat{\mathbf{w}}^t - \boldsymbol{\Lambda}_{:,n}(\hat{\underline{x}}_n^t - \hat{\underline{x}}_n^{t-1})$.
8: End for n.
9: $-$ Decoding step: For $n = 1, \ldots, N_q$, $\{\hat{\underline{\tau}}_n^t, v_{\underline{\tau}_n}^t\} \rightarrow$ Decoder.
10: Compute (2.5) by using the extrinsic information from the decoder.
11: $-$ Denoising step: For $n = 1, \ldots, N_q$, compute $\{\hat{\underline{x}}_n^t, v_{\underline{x}_n}^t\}$ from (2.44); $\hat{\mathbf{x}}^t = [\hat{\underline{x}}_1^t, \ldots, \hat{\underline{x}}_{N_q}^t]^T$.

2.4.2 DCT Model Based Parallel BP-MF Message Passing Schedule

Given $\underline{b}(x_i)$ and $b(\boldsymbol{\alpha})$ from the previous iteration, the message $m_{f_{r_n} \rightarrow x_i}(x_i)$ is formulated based on the MF method by

$$m_{f_{r_n} \rightarrow x_i}(x_i) \propto \exp\left\{ E_{b(\boldsymbol{\alpha}) \prod_{j \neq i} \underline{b}(x_j)} \left[\log f_{r_n}(\mathbf{x}, \boldsymbol{\alpha}) \right] \right\} \propto \mathcal{N}_c(x_i; \hat{\tau}_i^{'}, v_{\tau_i^{'}}), \tag{2.50}$$

where we have:

$$v_{\tau_i^{'}} = 2\sigma^2 / (\hat{\varrho}_{u_n} |\Lambda_{n,i}|^2),$$

$$\hat{\tau}_i^{'} = \frac{v_{\tau_i^{'}} \hat{\varrho}_{u_n} \Lambda_{n,i}^*}{2\sigma^2} \left[(\mathbf{\Phi}\hat{\boldsymbol{\alpha}})_n^* r_n / \hat{\varrho}_{u_n} - \mathbf{\Lambda}_{n,:} \hat{\mathbf{x}}_{\bar{i}} \right], \tag{2.51}$$

and the message $m_{x_i \to f_{M,g}}^{\mathrm{BP-MF}}(x_i) \propto \mathcal{N}_c(x_i; \underline{\hat{\tau}}_i, v_{\underline{\tau}_i})$ is associated with

$$\begin{aligned} v_{\underline{\tau}_i} &= 2\sigma^2 \left(\mathbf{\Lambda}_{:,i}^H \mathcal{D}(\hat{\boldsymbol{\varrho}}_u) \mathbf{\Lambda}_{:,i} \right)^{-1}, \\ \underline{\hat{\tau}}_i &= \frac{v_{\underline{\tau}_i}}{2\sigma^2} \mathbf{\Lambda}_{:,i}^H \mathcal{D}(\hat{\boldsymbol{\varrho}}_u) \left(\mathcal{D}^H(\mathbf{\Phi}\hat{\boldsymbol{\alpha}}) \mathcal{D}^{-1}(\hat{\boldsymbol{\varrho}}_u) \mathbf{r} - \mathbf{\Lambda}\hat{\mathbf{x}}_{\bar{i}} \right), \end{aligned} \tag{2.52}$$

where $\hat{\boldsymbol{\varrho}}_u = [\hat{\varrho}_{u_1}, \ldots, \hat{\varrho}_{u_{N_q}}]^T$.

The message emerging from the observation node to the PHN estimation subgraph can be expressed as:

$$\begin{aligned} m_{f_{r_n} \to \boldsymbol{\alpha}}(\boldsymbol{\alpha}) &\propto \exp\left\{ E_{\underline{b}(\mathbf{x})} \left[\log f_{r_n(\boldsymbol{\alpha}, \mathbf{x})} \right] \right\} \\ &\propto \exp\left\{ \frac{2\Re\{ r_n^* \mathbf{\Lambda}_{n,:} \hat{\mathbf{x}} (\mathbf{\Phi}\boldsymbol{\alpha})_n \} - |(\mathbf{\Phi}\boldsymbol{\alpha})_n|^2 \hat{\varrho}_{x_n}}{2\sigma^2} \right\}, \end{aligned}$$

where $\hat{\varrho}_{x_n} = \sum_{i=1}^{N_q} |\Lambda_{n,i}|^2 (|\hat{x}_i|^2 + v_{x_i}) + 2 \sum_{i=1}^{N_q} \sum_{\jmath > i}^{N_q} \Re\left\{ \Lambda_{n,i}^* \Lambda_{n,\jmath} \hat{x}_i^* \hat{x}_\jmath \right\}$. Thus, the *a posteriori* pdf of the DCT coefficient vector is given by

$$\begin{aligned} b(\boldsymbol{\alpha}) &\propto p(\boldsymbol{\alpha}) \prod_{n=1}^{N_q} m_{f_{r_n} \to \boldsymbol{\alpha}}(\boldsymbol{\alpha}) \\ &\propto \exp\left\{ \frac{2\Re\{ (\mathbf{r} \odot (\mathbf{\Lambda}\hat{\mathbf{x}})^*)^H \mathbf{\Phi}\boldsymbol{\alpha} \} - \boldsymbol{\alpha}^H \mathbf{\Phi}^H \mathcal{D}(\hat{\boldsymbol{\varrho}}_x) \mathbf{\Phi}\boldsymbol{\alpha}}{2\sigma^2} \right\}, \end{aligned}$$

where $\hat{\boldsymbol{\varrho}}_x = [\hat{\varrho}_{x_1}, \ldots, \hat{\varrho}_{x_{N_q}}]^T$, and its covariance matrix and mean vector are approximated as

$$\mathbf{V}_{\boldsymbol{\alpha}} = 2\sigma^2 / \bar{\varrho}_x \mathbf{I}_{L_{\mathrm{PN}}}, \quad \hat{\boldsymbol{\alpha}} = \frac{1}{\bar{\varrho}_x} \mathbf{\Phi}^H (\mathbf{r} \odot (\mathbf{\Lambda}\hat{\mathbf{x}})^*), \tag{2.53}$$

where $\bar{\varrho}_x$ is the average of the elements in $\hat{\boldsymbol{\varrho}}_x$.

The amalgamated BP-MF using the parallel message passing schedule for the DCT model based joint PHN estimation and decoding algorithm, which is referred to as BP-MF-p, is summarized in Algorithm 4.

2.4.3 Complexity Analysis

In Table 1, we provide a brief complexity comparison of the proposed GAMP-Wiener, GAMP-DCT, BP-MF-s and BP-MF-p algorithms. All of these methods involve the calculations of the standard BCJR decoding, thus we only analyze the

Algorithm 4 Amalgamated BP-MF algorithm for joint PHN estimation and decoding based on the DCT model using the parallel scheduling.

1: Initialization: $t = 0$ (iteration index)

2: For $n = 1, \ldots, N_q$, compute $\hat{x}_n^0$ and $v_{x_n}^0$.

3: Set $t = 1$ and repeat the following steps until $\sum_n |\hat{x}_n^{t+1} - \hat{x}_n^t|^2 < \epsilon \sum_n |\hat{x}_n^t|^2$ or $t > T_{\max}$.

4: $-$ PHN estimation step: Compute $\{\mathbf{V}_{\boldsymbol{\alpha}}^t, \hat{\boldsymbol{\alpha}}^t\}$ from (2.53).

5: $-$ Soft demodulation step: Compute $\hat{\mathbf{w}}^t = \mathcal{D}^H(\boldsymbol{\Phi}\hat{\boldsymbol{\alpha}}^t)\mathcal{D}^{-1}(\hat{\boldsymbol{\varrho}}_u^t)\mathbf{r} - \boldsymbol{\Lambda}\hat{\mathbf{x}}^{t-1}$.

6: For $n = 1, \ldots, N_q$, compute $v_{\underline{\tau}_n}^t$ from (2.52); compute $\hat{\underline{\tau}}_n^t = \frac{v_{\underline{\tau}_n}}{2\sigma^2}\boldsymbol{\Lambda}_{:,n}^H \mathcal{D}(\hat{\boldsymbol{\varrho}}_u^t)\hat{\mathbf{w}}^t + \hat{x}_n^{t-1}$.

7: $-$ Decoding step: For $n = 1, \ldots, N_q$, $\{\hat{\underline{\tau}}_n^t, v_{\underline{\tau}_n}^t\} \rightarrow$ Decoder.

8: Compute (2.5) by using the extrinsic information from the decoder.

9: $-$ Denoising step: For $n = 1, \ldots, N_q$, compute $\{\hat{\underline{x}}_n^t, v_{\underline{x}_n}^t\}$ from (2.44); $\hat{\mathbf{x}}^t = [\hat{\underline{x}}_1^t, \ldots, \hat{\underline{x}}_{N_q}^t]^T$.

Table 2.1: Comparisons of computational complexity (For example: $N_q = 64$, $L_{PN} = 2$, $G = 16$, $S = 4$, and $M_g = 4$.)

Algorithm	Complexity of the demodulator	Value	Complexity of the PHN estimator	Value
GAMP-Wiener	$\mathcal{O}(GN_gS + 2N_q \log N_q)$	$\mathcal{O}(1024)$	$\mathcal{O}(N_q)$	$\mathcal{O}(64)$
GAMP-DCT	$\mathcal{O}(GN_gS + 2N_q \log N_q)$	$\mathcal{O}(1024)$	$\mathcal{O}(2N_qL_{PN})$	$\mathcal{O}(256)$
BP-MF-s	$\mathcal{O}(GN_gS + N_q \log N_q + N_q^2)$	$\mathcal{O}(4736)$	$\mathcal{O}(N_q)$	$\mathcal{O}(64)$
BP-MF-p	$\mathcal{O}(GN_gS + 2N_q \log N_q)$	$\mathcal{O}(1024)$	$\mathcal{O}(2N_qL_{PN})$	$\mathcal{O}(256)$
Optimal Detector	$\mathcal{O}(2^{GM_g})$	$\mathcal{O}(1.8 \times 10^{19})$	–	–
EKF [37]	–	–	$\mathcal{O}(N_q)$	$\mathcal{O}(64)$
VI [38]	–	–	$\mathcal{O}(N_q^3)$	$\mathcal{O}(2.6 \times 10^5)$

computational complexity of a single iteration between the soft demodulator and PHN estimator within an OFDM-IM block. For soft demodulation by the GAMP-Wiener scheme, the order of $\mathcal{O}(GN_gS)$ operations are needed for the calculation of (2.11) and (2.12), plus an N_q-point IFFT for evaluating (2.24) and an N_q-point FFT for evaluating (2.26). We point out that the complexity in this part grows linearly both with the number of IM symbols and the modulation order, while that in the optimal receiver has a complexity order of $\mathcal{O}(2^{GM_g})$. The PHN estimation, i.e. the computation of (2.28) and the messages in the PHN estimation

subgraph based on the Wiener model and on the second-order Taylor expansion approximation is in the order of $\mathcal{O}(N_q)$. The GAMP-Wiener and GAMP-DCT scheme differ in the PHN estimation part. Explicitly, in GAMP-DCT, $\mathcal{O}(2N_qL_{\mathrm{PN}})$ operations are required for the calculation of (2.39) and (2.41) following from the assumption of $\mathcal{D}(\hat{\boldsymbol{\varrho}}_z) \approx \hat{\varrho}_z\mathbf{I}_{N_q}$. The BP-MF-s differs from the GAMP-Wiener scheme in the soft demodulation, where the complexity is increased from $\mathcal{O}(N_q \log N_q)$ to $\mathcal{O}(N_q^2)$, since the FFT cannot be directly invoked, when the messages are updated sequentially. By contrast, the BP-MF-p and GAMP-DCT have similar complexity in this part using the FFT, when passing messages in parallel. We also compare the complexity of the proposed algorithms to that of the conventional pilot based PHN estimation method of [82] combined with the conventional OFDM-IM detector of [13]. The calculations of a frequency-domain received sequence need an N_q-point FFT. The PHN estimation part in the pilot based method requires $\mathcal{O}(N_{\mathrm{pilot}})$ operations, where N_{pilot} is the total number of pilots contained within an OFDM-IM packet. However, the complexity of these calculations is modest, which have already been performed before starting iterative decoding. After compensating the received samples using the PHN estimates, the conventional LLR detector of [13] is employed for soft demodulation, which has a complexity order of $\mathcal{O}(GM_gS)$. Furthermore, the existing PHN tracking algorithms of [37] and [38] are combined with the proposed demodulation method and the complexity of PHN estimation is studied for comparison with both its Wiener model and the DCT model based counterparts. For the EKF of [37], similar PHN estimation complexity is seen as that in the GAMP-Wiener scheme due to the local linearization of the system model using first-order Taylor expansion. The VI approach of [38] involves an inversion of an $(N_q \times N_q)$-element matrix, which imposes a complexity order of $\mathcal{O}(N_q^3)$. In Table. 1, we give an example for $N_q = 64$, $L_{PN} = 2$, $G = 16$, $S = 4$, and $M_g = 4$. We observe that the proposed PHN estimation algorithms are computationally more efficient compared to the VI approach, for example, by a factor of 4096 at $N_q = 64$, when we employ the GAMP-Wiener method.

2.5 Simulation Results

In this section, simulation results are provided for evaluating the performance of the proposed joint PHN estimation and decoding algorithms for the OFDM-IM scheme and for comparing it with conventional OFDM. For all simulations, a rate-0.66 (i.e., $R_c = 0.66$) LDPC code[d] using a block size 8064 is used. The maximum number of iterations in the LDPC decoder is 50, while it is $T_{\max} = 20$ between the SISO demodulation and SISO decoding, unless otherwise stated. The tolerance value is set to be $\epsilon = 10^{-12}$. The number of subcarriers in an OFDM-IM symbol is kept constant at 64, i.e., $N_q = 64$. A ten-tap (i.e., $L_{\rm h} = 10$) Rayleigh fading channel is considered. The CIRs are circularly symmetric complex Gaussian random variables following the distribution $\mathcal{N}(h_l; 0, 1/L_{\rm h})$. The CP length $L_{\rm CP}$ is selected to be 13. The SNR is expressed as E_b/N_0, with $E_b = \frac{N_q+L_{\rm CP}}{R_c G(M_{g1}+M_{g2})}$ being the energy per bit. The effective throughput is $R_c \frac{G(M_{g1}+M_{g2})}{N_q+L_{\rm CP}}$ [bits/symbol]. For characterizing the PHN, a Wiener model with $\sigma_\triangle = 0.5^\circ$ and a severer one with $\sigma_\triangle = 3^\circ$ are used for generating PHN samples. The initial phase θ_1 is uniformly distributed in $[0, 2\pi)$. The number of DCT coefficients $L_{\rm PN}$ is designed to be 2, unless otherwise specified. A pilot is inserted every 64 subcarriers to assist the iterative decoder to bootstrap and to aid synchronization[e]. Bearing in mind that OFDM-IM scheme tends to exhibit benefits over its OFDM counterpart when the throughput is below 2 [bits/symbol] [34], we only consider low-order modulation schemes[f] such as binary phase shift keying (BPSK), i.e., $S = 2$, and quadrature phase shift keying (QPSK), i.e., $S = 4$.

Fig. 2.4 shows the BERs of the proposed GAMP-Wiener and GAMP-DCT algorithms for different OFDM-IM schemes characterized by $\{S = 4, N_g = 4, K_g =$

[d]The variable and check node degree distributions (refer to [72] for details) are $X_1 = 0.00003 + 0.1909X + 0.4070X^2 + 0.4020X^7$ and $X_2 = 0.5029X^9 + 0.4971X^{10}$, respectively. The parameter definitions follow [72].

[e]The power overhead due to the pilots is included in the simulation results. The gap between cases with known PHN and unknown PHN only comes from different algorithms and is not due to pilots, since the increase of required SNR is added artificially in known PHN cases.

[f]As evidenced by our additional extensive simulations, the BER performance advantage of the OFDM-IM scheme over its classic OFDM counterpart tends to erode for higher order modulations.

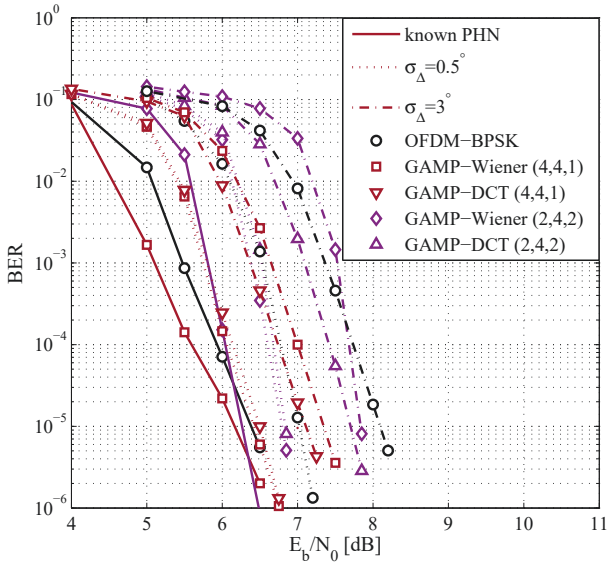

Figure 2.4: BER performance of classic BPSK-modulated OFDM and OFDM-IM (characterized by (4,4,1) and (2,4,2)) systems communicating over a 10-tap Rayleigh fading channel affected by PHN. Wiener PHN model with standard deviations of $\sigma_{\triangle} = 0.5°$ and $3°$ are considered. Both proposed GAMP-Wiener and GAMP-DCT are evaluated. The effective throughput is 0.55 bits/symbol.

1} (denoted by "$(4, 4, 1)$") and $\{S = 2, N_g = 4, K_g = 2\}$ (denoted by "$(2, 4, 2)$"), respectively. For comparison, the BER curves of classic BPSK modulated OFDM (denoted by "OFDM BPSK"), yielding the same transmission rate as that of OFDM-IM, are also plotted. The BERs of the OFDM-IM and OFDM schemes with known PHN, denoted by "known PHN", are also shown as benchmarks. A constellation rotation technique with a rotation angle of $\pi/12$ and the coordinate interleaved orthogonal design (CIOD) approach of [36] are used for improving the diversity gain. Additionally, for OFDM-IM with known PHN, the subcarrier-level interleaving scheme of [35] is also invoked for attaining multipath diversity by reducing the correlation amongst the subcarriers 1 to N_q. As shown in Fig. 2.4, in case of known PHN, OFDM-IM outperforms its OFDM counterpart in the high-SNR region, in agreement with the findings of [13,34,35]. This is because the performance advantage of the bits transmitted in the IM domain is more significant at high SNRs, especially for $\sigma_{\triangle} = 0.5°$. Specifically, the performance loss of both the GAMP-Wiener and GAMP-DCT algorithms recorded in case of known PHN is modest, which confirms the efficiency of the Wiener model and DCT model based PHN estimation methods,

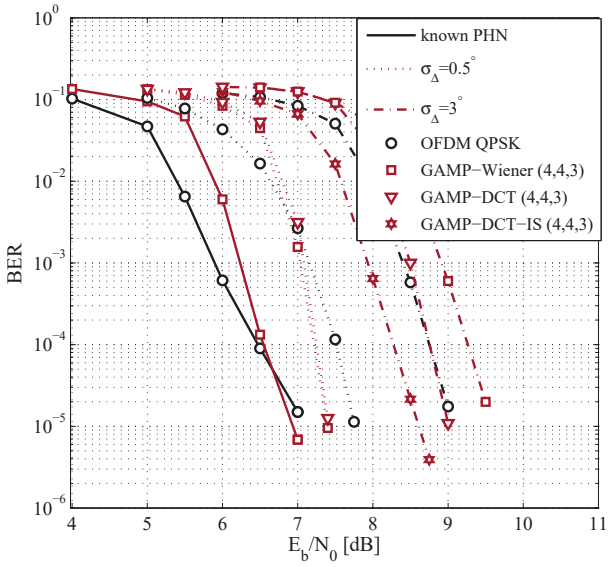

Figure 2.5: BER performance of classic QPSK-modulated OFDM and OFDM-IM (characterized by (4,4,3)) systems communicating over a 10-tap Rayleigh fading channel affected by PHN. Wiener PHN model with standard deviations of $\sigma_{\triangle} = 0.5°$ and $3°$ are considered. Both proposed GAMP-Wiener and GAMP-DCT are evaluated. The effective throughput is 1.1 bits/symbol.

when $\sigma_{\triangle} = 0.5°$. By contrast, about 0.7 dB gap is seen for classic OFDM, if the target BER is 10^{-5}. This means that using the proposed GAMP based algorithms, the OFDM-IM schemes transmitting part of the information bits by the indices of activated subcarriers indeed succeed in improving the performance over classic OFDM with PHN distortion. For $\sigma_{\triangle} = 3°$, the GAMP-DCT is slightly superior to GAMP-Wiener for OFDM-IM $(4, 4, 1)$ at a slightly higher complexity. This indicates that the accuracy of the Taylor expansion in the GAMP-Wiener receiver degrades upon increasing the PHN intensity. Meanwhile, we point out that the correlation between the PHN samples is not exploited across different OFDM-IM symbols, which is also a reason for the performance degradation of GAMP-Wiener. Explicitly, the mean values of the messages $m_{f_{\theta_n}\to\theta_n}(\theta_n)$ and $m_{f_{\theta_{n+1}}\to\theta_n}(\theta_n)$ at the boundaries of an OFDM-IM symbol are initialized with zeros, which will introduce a transient for the PHN estimation, especially when N_q is not long enough.

In Fig. 2.5, we compare the BER performance of the OFDM-IM $(4, 4, 3)$ and classic QPSK modulated OFDM both at an effective throughput of 1.1 bits/symbol. As seen from Fig. 2.5, the proposed GAMP-Wiener achieves better BER perfor-

mance than classic OFDM beyond a certain SNR when $\sigma_{\triangle} = 0.5°$. This trend is similar to that in Fig. 2.4. However, this performance advantage diminishes for a relatively severe PHN of $\sigma_{\triangle} = 3°$. This can be explained by the fact that the accuracy of data detection and PHN estimation is reduced upon increasing the PHN intensity, since more ICI is introduced. To improve the BER performance of OFDM-IM $(4, 4, 3)$ for $\sigma_{\triangle} = 3°$, we add a depth-N_g subcarrier interleaver to rearrange the contaminated subcarriers in frequency domain. The BER curve of this interleaving aided GAMP-DCT method is labelled by "GAMP-DCT-IS". Specifically, the interleaved frequency domain signals can be expressed as $\breve{\mathbf{x}} = [x_1(1), x_2(1), \ldots, x_G(1), \ldots, x_1(N_g), \ldots, x_G(N_g)]^T$. Accordingly, after de-interleaving at the receiver, we have

$$\mathbf{r} = \frac{1}{\sqrt{K_q}} \mathcal{D}(e^{j\boldsymbol{\theta}}) \breve{\mathbf{F}}^H \breve{\mathbf{H}} \mathbf{x} + \mathbf{w}, \tag{2.54}$$

where $\breve{\mathbf{F}}^H$ and $\breve{\mathbf{H}}$ are de-interleaved versions of $\mathbf{F}^H$ and $\mathbf{H}$, respectively. In this way, the performance degradation of IM-modulated bits due to the frequency-domain correlation is reduced. It is shown that as expected, about 0.4 dB gain is achieved at a BER of 10^{-5} for the GAMP-DCT-IS over GAMP-DCT. However, the OFDM-IM having realistic unknown PHN and using the interleaving loses the low complexity advantage of the FFT, which is in contrast to the known PHN case due to the matrix $\breve{\mathbf{F}}^H$ in (2.54). Using Equations (2.15), (2.16), (2.21) and (2.22) instead of (2.23), (2.24), (2.25) and (2.26), the complexity is increased from $\mathcal{O}(N_q \log N_q)$ to $\mathcal{O}(N_q^2)$. Hence, there is a flexible performance vs. complexity tradeoff in the GAMP based OFDM-IM receiver.

In Fig. 2.6, the BER performance of the BP-MF-s and BP-MF-p based receivers is compared to that of the proposed GAMP-Wiener and GAMP-DCT algorithms at a PHN of $\sigma_{\triangle} = 3°$. The OFDM-IM schemes characterized by $(4, 4, 1)$ and $(4, 4, 3)$ and having throughputs of 0.55 bits/symbol and 1.1 bits/symbol are considered. The number of turbo iterations between SISO demodulation and SISO decoding is fixed to $T_{\max} = 20$. We observe that GAMP-DCT exhibits the best BER amongst the aforementioned algorithms. Compared with GAMP-Wiener, BP-MF-s achieves slightly lower BER values, which is attributed to exploiting the *a posteriori* estimates

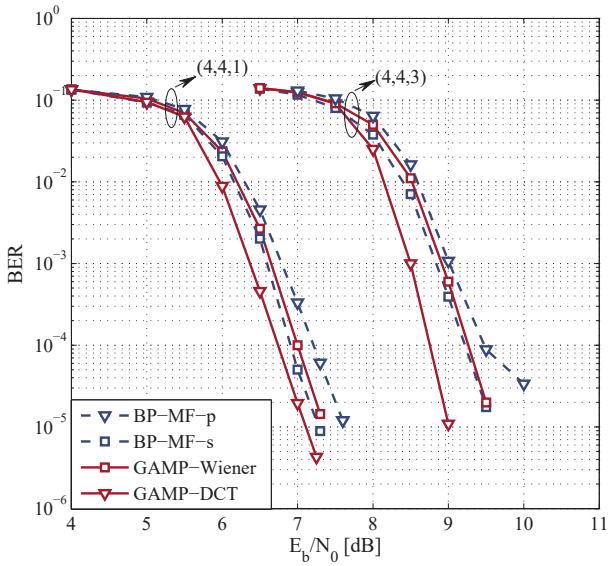

Figure 2.6: BER performance of the proposed GAMP-DCT, GAMP-Wiener and BP-MF algorithms. Both parallel and sequential schedules for the BP-MF methods are studied for comparison. Wiener PHN model with $\sigma_{\triangle} = 3^{\circ}$ is considered. The effective throughput of OFDM-IM $(4, 4, 1)$ and $(4, 4, 3)$ is 0.55 bits/symbol and 1.1 bits/symbol, respectively.

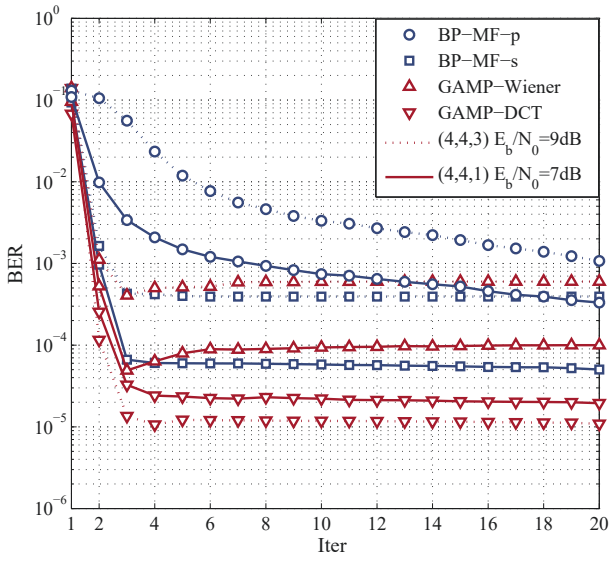

Figure 2.7: BER performance of the proposed GAMP-DCT, GAMP-Wiener and BP-MF algorithms versus iteration index. Both parallel and sequential schedules for the BP-MF methods are studied for comparison. Wiener PHN model with $\sigma_{\triangle} = 3^{\circ}$ is considered. The effective throughputs of OFDM-IM $(4, 4, 1)$ and $(4, 4, 3)$ are 0.55 bits/symbol and 1.1 bits/symbol, respectively.

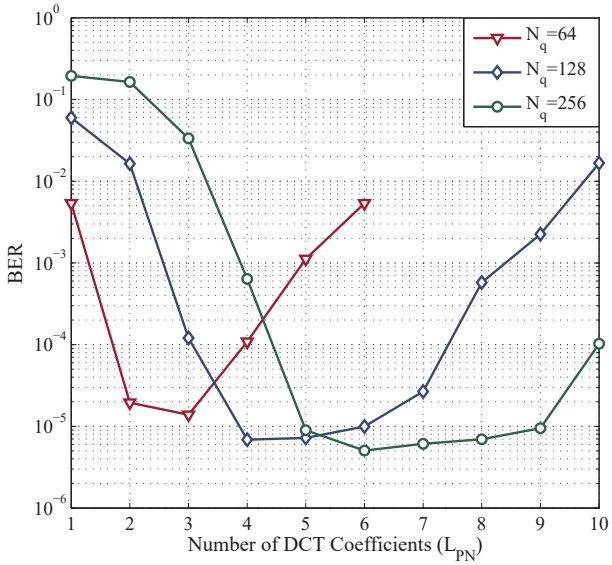

Figure 2.8: The impact of the number of DCT coefficients on BER performance of the GAMP-DCT algorithm at $E_b/N_0 = 7$ dB. The block length of DCT model is fixed at 64, 128, and 256, respectively. Wiener PHN model with $\sigma_{\triangle} = 3^{\circ}$ is considered.

of $\mathbf{x}_g$ in the estimation of the remaining symbols during iteration t. By contrast, a significant performance loss is seen for the BP-MF-p method, especially for OFDM-IM $(4, 4, 3)$. For example, at a BER of 10^{-5}, about 0.6 dB loss is observed for BP-MF-p over GAMP-DCT for OFDM-IM $(4, 4, 1)$, but an error floor around the BER value of 10^{-4} for OFDM-IM $(4, 4, 3)$. Furthermore, Fig. 2.7 plots the BERs of the receivers as a function of the iteration index under the conditions of Fig. 2.6. The SNR is fixed to 7 dB (dotted line) and 9 dB (cont. line) for OFDM-IM $(4, 4, 1)$ and OFDM-IM $(4, 4, 3)$, respectively. Observe that the GAMP-Wiener, GAMP-DCT and BP-MF-s algorithms all converge after about 5 iterations. Therefore, 5 iterations on average are sufficient to meet the pre-specified stopping criterion. By contrast, the convergence of BP-MF-p is slow compared to both its sequential counterpart and to the GAMP based receivers, which explains its poor BER performance at high SNRs in Fig. 2.6. We also observe that although the complexity of GAMP-Wiener is lower than that of BP-MF-s and its performance is inferior to BP-MF-s after convergence, the BER of GAMP-Wiener is slightly lower than that of BP-MF-s at the first few iterations.

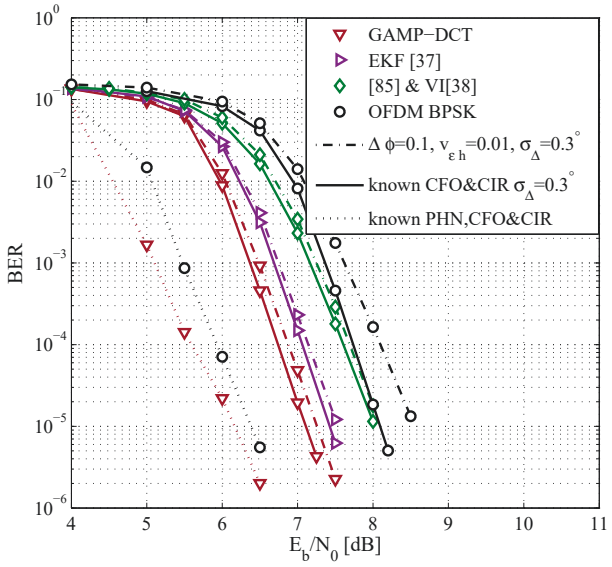

Figure 2.9: The impact of residual CFO of $\Delta\phi = 0.1$ and imperfect CSI with parameter $v_{\epsilon_h} = 0.01$ on BER performance of the GAMP-DCT algorithm, the existing EKF [37], and the [85] combined with VI [38]. Wiener PHN model with $\sigma_{\triangle} = 3^{\circ}$ is considered.

In Fig. 2.8, the BER performance vs. the number of DCT coefficients is studied. Observing that similar to the phenomenon shown in [71] for a single-carrier system, given the OFDM-IM symbol length, there is an appropriate number of DCT coefficients. After optimizing the relationship between the number of DCT coefficients and OFDM-IM symbol length, we find that a slightly better BER performance can be attained for higher N_q.

The proposed algorithm relying on residual CFO $\Delta\phi$ and imperfect channel information is evaluated in Fig. 2.9. In practical systems, the channel estimation error at the receiver is assumed to be a zero-mean Gaussian variable with a variance of v_{ϵ_h} [83]. Fig. 2.9 compares the BER performance of the proposed GAMP-DCT both for OFDM-IM and for classic OFDM, when $\Delta\phi = 0.1$ [84], $v_{\epsilon_h} = 0.01$ and $\sigma_{\triangle} = 3^{\circ}$. The simulation settings are kept the same as in Fig. 2.4. In order to take the channel uncertainty and the residual CFO into account, the proposed algorithm is modified by recalculating the equivalent mean and power of the residual interference imposed by residual CFO and imperfect channel estimation using the PHN estimates and symbols estimates at each iteration. Additionally, the BER curves of the pilot

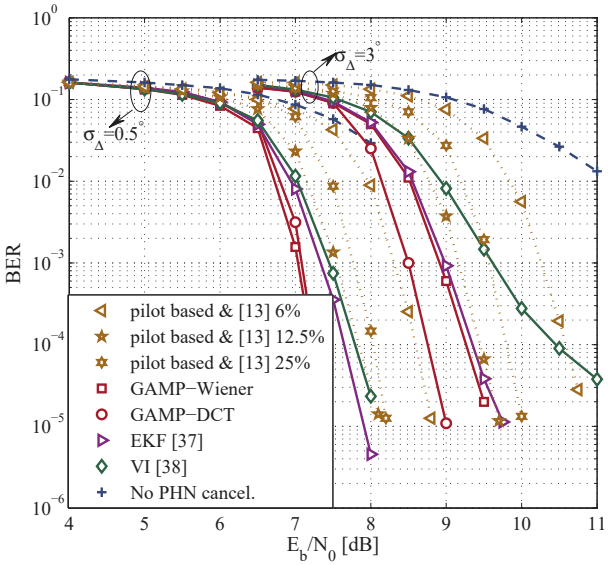

Figure 2.10: The BER performance of proposed GAMP-Wiener and GAMP-DCT compared with existing EKF [37] and VI [38] algorithms. The OFDM-IM is characterized by $(4, 4, 3)$. Wiener PHN model with $\sigma_\triangle = 0.5°$ and $3°$ are considered. The effective throughput is 1.1 bits/symbol.

based joint CFO, CIR and PHN estimation methods of [37] (denoted by "EKF [37]") and of [85] are now combined with the PHN tracking algorithm of [38] (denoted by " [85] & VI [38]"), which are shown for comparison. We observe that the proposed GAMP-DCT method outperforms the existing pilot based EKF and [85] & VI, even in the presence of residual CFO and channel estimation errors. Moreover, compared to classic OFDM, OFDM-IM relying on the proposed GAMP-DCT method is more robust both to the residual CFO and to the channel uncertainty. For example, for $\Delta\phi = 0.1$, $v_{\epsilon_h} = 0.01$, and $\sigma_\triangle = 3°$, only a modest performance gap of 0.15 dB is seen for OFDM-IM compared to the perfect CFO and CSI based BER curve at BER$= 10^{-5}$, while 0.4 dB is observed for classic OFDM. However, this gap widens at higher SNR values.

In Fig. 2.10, the BER performance of the proposed GAMP-Wiener and GAMP-DCT algorithms is compared to existing methods. The conventional pilot based PHN estimation method of [82] is now combined with the conventional OFDM-IM detector of [13] (denoted by "pilot based & [13]"). We consider different ratios of pilots, e.g., 6%, 12.5% and 25%. The extended Kalman filter method (denoted by

"EKF [37]") and the variational inference approach (denoted by "VI [38]") were proposed for joint PHN estimation and uncoded ODFM data detection in [37] and [38], respectively. These are now extended to the coded OFDM-IM scheme by combining it with the proposed GAMP based data detection algorithm. The system parameters are the same as in Fig. 2.5, where two different PHN variances, i.e., $\sigma_{\triangle} = 0.5^{\circ}$ and 3° are considered. Additionally, the BER performance recorded without PHN estimation (denoted by "No PHN cancel.") is also shown for comparison. It is seen that the OFDM-IM performance degrades significantly, if the effect of PHN is ignored. The conventional pilot based method improves the BER performance by increasing the ratio of pilots from 6% to 12.5%. However, a performance erosion is imposed by further increasing this ratio, since due to the power assigned to the pilots, the E_b/N_0 reduction becomes significant. We also observe that the pilot based algorithm is inferior to other data-aided methods, due to the fact that the PHN varies rapidly compared the OFDM-IM symbol rate and hence the PHN cannot be sufficiently mitigated using pilots. By using the proposed GAMP-Wiener and GAMP-DCT, the BER performance improves considerably even for severe PHN. The GAMP-DCT scheme outperforms the others due to the fact that the first L_{PN} DCT coefficients are well suited for the compact representation of the lowpass process of discrete-time PHN samples [71][g]. However, the VI method exhibits poor BER performance compared to the proposed ones and the EKF, especially at high SNRs, where the performance is mainly limited by the PHN estimation accuracy instead of the additive white noise. Explicitly, it suffers from an error floor when $\sigma_{\triangle} = 3^{\circ}$. This outcome can be attributed to the first-order Taylor expansion approximation of the PHN vector over the whole OFDM-IM symbol. The linearization approximation directly imposed on the system model substantially deviates from the true system parameters in the presence of strong PHN. By contrast, both the proposed GAMP-Wiener method and the EKF updates the current PHN sample estimation by using previous PHN estimates. The essential difference between the EKF and the GAMP-Wiener is that the EKF linearizes the system model locally using the first-order

[g]Similar applications of DCT expansion tehcnique can be found in various audio and image compression algorithms.

Taylor series, while the GAMP-Wiener approximates the messages exchanged in the factor graph by second-order Taylor expansion. This explains the superiority of the GAMP-Wiener over the EKF for two different PHN variance values.

2.6 Conclusion

In this chapter, low-complexity joint PHN estimation and decoding methods are developed in the framework of message passing on a factor graph. Both the Wiener process and the truncated DCT expansion model are considered for approximating the PHN variation. Then based on these a factor graph is constructed for explicitly representing the joint estimation and detection problem. Taking full account of the sparse and structured *a priori* information arriving from the SISO decoder of a turbo receiver, a modified GAMP algorithm is invoked for decoupling the frequency-domain symbols. In the decoupling step, MF approximation is employed for solving the unknown nonlinear transform matrix problem imposed by PHN. Furthermore, merged BP and MF methods amalgamated both with sequential and parallel message passing schedules are introduced and compared to the proposed GAMP based algorithms in terms of their BER vs. complexity. Our simulation results demonstrate the efficiency of the proposed algorithms in the presence of both perfect and imperfect channel state information.

Chapter 3

Iterative Detection for Dual-Mode-Aided IM in Nonlinear Dispersive Satellite Channels

In the previous chapter, multi-carrier OFDM-IM systems are studied, which however are sensitive to nonlinear distortions due to their high PAPR. In this chapter, the DM-aided IM technique is used to improve the spectral efficiency of single-carrier Satcom and then two low-complexity iterative detection algorithms are developed for jointly detecting IM symbols and conventional constellation symbols while eliminating the effects of nonlinear inter-symbol interference.

3.1 Introduction

The integration between terrestrial communications and Satcom has been a new trend for satisfying the challenging requirements of high performance and seamless services. However, both bandwidth and available power are limited over satellite channels. In order to keep up with the rapid growth of terrestrial communications, Satcom should exploit innovative techniques to improve its spectrum utilization and energy efficiency. The spectral efficiency can be improved by using high-order modulation. However, higher-order modulation is more sensitive to nonlinear channels. By using the DM-aided IM technique, the specific permutation of the

two distinguished modes is used for implicitly conveying extra information, hence improving the achievable throughput at a given BER performance. In contrast to existing IM-based terrestrial communication systems, which are mainly formulated in linear frequency-flat [12] or linear frequency-selective Rayleigh fading channels [13, 14, 40], Satcom generally suffers from nonlinear and dispersive distortions[a] [41]. Consequently, the receiver has to combat ISI and nonlinear distortions, as well as to jointly detect the IM bits and the ordinary constellation mapping bits. From existing literature, iterative detection methods for IM in nonlinear dispersive satellite channels have not been studied.

In Satcom without using IM, many algorithms have been developed for SC signal detection [86–90]. SC modulation has low PAPR and hence is less vulnerable to nonlinear distortions compared with OFDM[b] [23]. In [86], a low-complexity detector based on Gaussian approximation of the nonlinear ISI terms was developed for SC in Satcom. However, it delivers remarkably degraded performance compared with the optimal receiver. In [87], the sum-product algorithm (SPA) was employed for detecting PSK symbols in the framework of factor graph (FG) [64]. The complexity of this detector increases linearly with the nonlinear channel memory at the cost of significant performance loss. By using the Markov chain Monte Carlo (MCMC) technique, a nonlinear equalizer was studied in [91], which offers an attractive tradeoff between the BER performance and the complexity.

Unfortunately, all existing detectors designed for conventional Satcom cannot be directly applied in Satcom relying on IM. One of the reasons is that not all permutations of the two distinguishable constellation modes[c] are utilized by the transmitter, and this has to be carefully taken into account at the receiver. Moreover, the discrete constellation symbols are priorly dependent due to the use of IM, which

[a] In a high-frequency band, nonlinear distortions are introduced by the power amplifier which is always driven near its saturation point in the satellite transponder or in the base station [23].

[b] Nonlinear distortions result in a loss of subcarrier orthogonality of multi-carrier OFDM systems, which inevitably leads to inter-channel interference. This means that the detector for OFDM-IM in nonlinear dispersive satellite channels has to combat ICI and nonlinear distortions.

[c] Single-mode-aided IM can be regarded as a special case of DMIM by setting the secondary constellation to be null.

is not the case in conventional Satcom. In this context, the widely-used independent Gaussian approximation of discrete symbol variables in original Satcom will lead to significant performance loss when directly used in the IM-aided Satcom without any modification. On the other hand, nonlinear ISI introduced by satellite channels results in a loss of index-modulated symbol vector (IM-symbol) orthogonality, i.e., the subframe orthogonality, which leads to the complexity of the optimal detector growing exponentially with the number of bits per frame [13].

Against this background, in this chapter, two low-complexity iterative detection algorithms are proposed for Satcom relying on DMIM (Sat-DMIM) in nonlinear dispersive satellite channels. The contributions of this chapter are summarized as follows:

- Instead of directly treating the correlated Sat-DMIM symbols as independent Gaussian random variables, the mode-selection pattern constraint is considered by explicitly calculating the *a priori* pdf of Sat-DMIM signal. Given this *a priori* information, the joint *a posteriori* pdf of discrete Sat-DMIM symbols is derived based on a Volterra-series expansion model [92] for nonlinear dispersive satellite channels.

- In order to eliminate the effects of both linear and nonlinear ISI over satellite channels, a suitably-designed factor graph is constructed by factorizing the joint *a posteriori* probability distribution, based on which an amalgamated BP and MF message passing algorithm is derived. Specifically, the MF method is applied at the observation nodes for decoupling the Sat-DMIM symbols, while the BP technique is invoked at the mode-selection pattern constraint nodes and at the PSK/QAM modulation constraint nodes for updating the estimates of IM bits and ordinary mapping bits.

- In contrast to directly approximating the nonlinear system model [87,93,94], a partial linearization approximation-aided BP-MF algorithm is proposed to further reduce the complexity. Specifically, by ignoring some high-order ISI terms, the message parameters to be updated with iterations are simplified. The number of sufficient statistics to be tracked for updating nonlinear messages

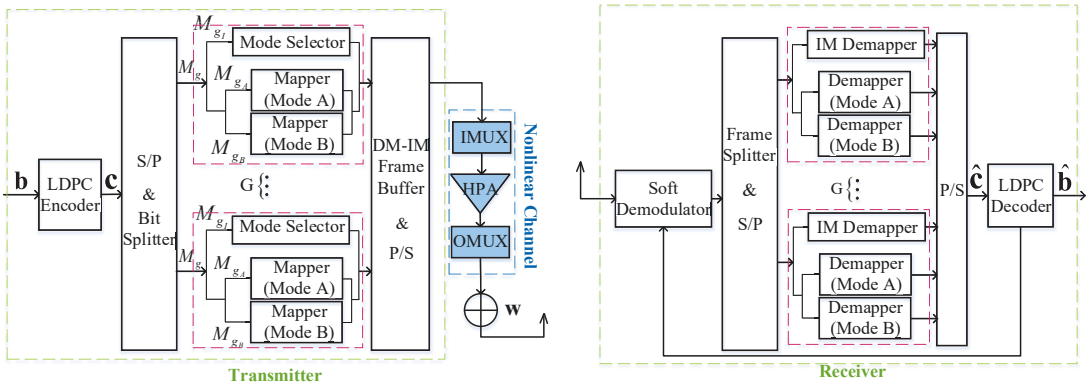

Figure 3.1: The transceiver of the LDPC-coded Satcom relying on DMIM.

is further reduced by using the first-order Taylor series expansion technique.

The rest of this chapter is outlined as follows. Section 3.2 introduces the system model. Low-complexity BP-MF-based and partial linearization approximation-based iterative detection algorithms are derived over a suitably-designed factor graph in Sections 3.3 and 3.4, respectively. Section 3.5 provides computer simulation results and complexity analysis. Finally, Section 3.6 concludes this chapter.

3.2 System Model

The transceiver architecture of low-density parity-check (LDPC)-coded [72] Satcom relying on DMIM is depicted in Fig. 3.1. For each Sat-DMIM frame, N-length data symbols $\mathbf{x} = [x_1, x_2, \ldots, x_N]^T$ are equally divided into G subframes, i.e., $\mathbf{x} \triangleq [\mathbf{x}_1^T, \ldots, \mathbf{x}_G^T]^T$ with each subframe containing $N_g \triangleq N/G$ constellation symbols. In contrast to conventional single-carrier Satcom, where only a single constellation mode is employed for all time slots, the N_g symbols in each Sat-DMIM subframe are drawn from two distinguishable constellation modes. Therefore, in Sat-DMIM, the information is conveyed both by conventional PSK/QAM symbols and the permutation of the two modes predesigned for each subframe. The details of the transmitter and the receiver structures, as well as the characteristics of nonlinear dispersive satellite channels are provided in the following subsections.

Table 3.1: The Look-up table of mode-selection mapping between the IM bits and the legitimate set for $N_g = 4$ and $K_g = 2$.

M_{g_I}	$\mathcal{I}_A$	$\mathbf{x}_g$
[0 0]	$\{1, 3\}$	$[s_g^A(1), s_g^B(1), s_g^A(2), s_g^B(2)]^T$
[0 1]	$\{1, 4\}$	$[s_g^A(1), s_g^B(1), s_g^B(2), s_g^A(2)]^T$
[1 0]	$\{2, 3\}$	$[s_g^B(1), s_g^A(1), s_g^A(2), s_g^B(2)]^T$
[1 1]	$\{2, 4\}$	$[s_g^B(1), s_g^A(1), s_g^B(2), s_g^A(2)]^T$

3.2.1 Transmitter Model

As shown in Fig. 3.1, a length-M_i sequence of information bits $\mathbf{b} \triangleq \{b_i\}_{i=1}^{M_i}$ is encoded by a rate-R_c LDPC encoder, and then the resultant $M_c \triangleq M_i/R_\mathrm{c}$ coded bits $\mathbf{c} = \{c_i\}_{i=1}^{M_c}$ are equally partitioned into G length-M_g ($M_g \triangleq M_c/G$) subgroups, i.e., $\mathbf{c} \triangleq [\mathbf{c}_1^T, \ldots, \mathbf{c}_G^T]^T$, $\mathbf{c}_g \in \{0,1\}^{M_g}$, $g = 1, \ldots, G$. By using the DMIM technique, the gth subgroup of coded bits $\mathbf{c}_g$ is mapped to the gth subframe $\mathbf{x}_g \triangleq [x_g(1), \ldots, x_g(N_g)]^T$, which is comprised of K_g symbols drawn from the P_A-ary constellation set $\dot{\mathcal{S}} = \{\dot{s}_1, \ldots, \dot{s}_{P_A}\}$ (denoted by "mode A") and the remaining $(N_g - K_g)$ symbols from the P_B-ary constellation set $\ddot{\mathcal{S}} = \{\ddot{s}_1, \ldots, \ddot{s}_{P_B}\}$ (denoted by "mode B"). Specifically, the M_g bits in $\mathbf{c}_g$ are further divided into smaller segments of M_{g_I}, M_{g_A} and M_{g_B} bits for different purposes: The first $M_{g_\mathrm{I}} = \left\lfloor \log_2 \binom{N_g}{K_g} \right\rfloor$ bits enter the mode selector to separate the indices of $\mathbf{x}_g$ into two index subsets $\mathcal{I}_g^A = \{i_{g,1}^A, \ldots, i_{g,K_g}^A\}$ and $\mathcal{I}_g^B = \{i_{g,1}^B, \ldots, i_{g,N_g-K_g}^B\}$, where $i_{g,m}^A, i_{g,l}^B \in \{1, \ldots, N_g\}$ for $m = 1, \ldots, K_g$ and $l = 1, \ldots, N_g - K_g$; the remaining $M_{g_A} = K_g \log_2 P_A$ and $M_{g_B} = (N_g - K_g) \log_2 P_B$ bits are mapped to K_g P_A-ary symbols $\mathbf{s}_g^A = [s_g^A(1), \ldots, s_g^A(K_g)]^T \in \dot{\mathcal{S}}^{K_g}$ and to $(N_g - K_g)$ P_B-ary symbols $\mathbf{s}_g^B = [s_g^B(1), \ldots, s_g^B(N_g - K_g)]^T \in \ddot{\mathcal{S}}^{(N_g - K_g)}$, respectively. Given the mode permutation chosen by the mode selector, the symbols in $\mathbf{s}_g^A$ and $\mathbf{s}_g^B$ are then assigned to the indices in $\mathcal{I}_g^A$ and $\mathcal{I}_g^B$, respectively, yielding the gth subframe $\mathbf{x}_g$. In other words, for $i = 1, \ldots, N_g$, we have $x_g(i) \in \dot{\mathcal{S}}$ for $i \in \mathcal{I}_g^A$, and $x_g(i) \in \ddot{\mathcal{S}}$ for $i \in \mathcal{I}_g^B$. Note that with a little abuse of notation, we may denote the coded bits corresponding to the vector $\mathbf{x}_g$ as $\mathbf{c}_g = [c_{g,1}^{\mathrm{IM}}, \ldots, c_{g,M_{g_\mathrm{I}}}^{\mathrm{IM}}, c_{g,1}^A, \ldots, c_{g,M_{g_A}}^A, c_{g,1}^B, \ldots, c_{g,M_{g_B}}^B]^T$, where $c_{g,i}^{\mathrm{IM}}$ is the ith IM bit; $c_{g,\jmath}^A$ and $c_{g,k}^B$ are the $\jmath$th and the kth ordinary constellation

mapping bits corresponding to the symbols belonging to $\dot{\mathcal{S}}$ and $\ddot{\mathcal{S}}$, respectively.

Let us refer to $\mathcal{I}_g^A$ as the mode-selection pattern, since the subset $\mathcal{I}_g^B$ is determined once $\mathcal{I}_g^A$ is known. Obviously, there are $\binom{N_g}{K_g}$ available permutations, while only $\Upsilon \triangleq 2^{M_{g_\mathrm{I}}}$ legitimate mode-selection patterns determined by the M_{g_I} IM bits are employed by the transmitter. Denote the predefined legitimate set as $\boldsymbol{\mathcal{I}}^A \triangleq \{\mathcal{I}^A(1), \ldots, \mathcal{I}^A(\Upsilon)\}$. The fact that $\mathcal{I}_g^A$ should belong to $\boldsymbol{\mathcal{I}}^A$ is referred to as the mode-selection pattern constraint. Considering that existing mode mapping methods tend to introduce error propagation if different modulation orders are employed [44–46], we use the improved mapping rule provided in Table 3.1 for $N_g = 4$ and $K_g = 2$ as an example. The legitimate set of mode-selection patterns in this example is $\boldsymbol{\mathcal{I}}^A = \{\{1,3\},\{1,4\},\{2,3\},\{2,4\}\}$. Specifically, according to the look-up Table 3.1, if the M_{g_I} IM bits of the mode selector are [0 0], the first and third entries of $\mathbf{x}_g$ are P_A-ary symbols associated with the following M_{g_A} bits, while the other two entries are P_B-ary symbols associated with the remaining M_{g_B} bits.

Finally, the frame buffer in Fig. 3.1 concatenates the G subframes $\{\mathbf{x}_g\}_{g=1}^G$, creating the length-N data frame $\mathbf{x} = [x_1, \ldots, x_N]^T$. The effective transmission rate of the Sat-DMIM scheme is given by

$$R = R_c \frac{\left\lfloor \log_2 \binom{N_g}{K_g} \right\rfloor + K_g \log_2 P_A + (N_g - K_g) \log_2 P_B}{N_g}.$$

3.2.2 Nonlinear Dispersive Satellite Channels

As shown in Fig. 3.1, the satellite transponder consists of an input demultiplexer (IMUX) filter, a high power amplifier (HPA) and an output multiplexer (OMUX) filter [95]. For reasons of power efficiency, the HPA always operates close to its saturation region, which inevitably introduces nonlinear memoryless distortion [88]. The IMUX and OMUX filters placed before and after HPA are used to reduce the effect of adjacent channels and to restrain the spectral broadening caused by the nonlinear amplifier, respectively. However, the employment of these two filters can introduce possibly serious ISI. We assume that the characteristics of the HPA, the IMUX and OMUX filters are perfectly known at both the transmitter and receiver sides. In order to describe the resultant effect of nonlinearities with memory, many

models are developed in literatures, including the polynomial model [96], as well as the Hammerstein model [97] and the Volterra model [92]. The Volterra model, derived by decomposing the nonlinear channels using the Taylor series, has a more generic expression. Based on the Volterra model, the nonlinear dispersive satellite channels can be formulated as a finite state machine, which has been widely applied for the receiver design in classical Satcom [87,89,90,98]. Note that taking more high-order Volterra series into consideration may be able to describe the actual satellite channels more accurately and help gain better performance of nonlinear detectors, but the computational complexity can become intractable and the instability of the nonlinear compensation process tends to increase [99]. Hence, we use a common Volterra-series expansion model of third order to decompose the nth discrete received signal as

$$r_n = \underbrace{\sum_{l=0}^{L} x_{n-l} h_l}_{\text{Linear ISI}} + \underbrace{\sum_{i=0}^{L} \sum_{j \geq i}^{L} \sum_{k=0}^{L} x_{n-i} x_{n-j} x_{n-k}^{*} h_{ijk}}_{\text{Nonlinear ISI}} + w_n, \tag{3.1}$$

where the parameter L describes the dispersion of the nonlinear channels; h_l and h_{ijk} are the first- and third-order Volterra kernels[d], respectively; and w_n is the zero-mean complex circularly symmetric Gaussian noise sample with a variance of $2\sigma^2$.

3.2.3 Receiver Model

In contrast to the conventional SC scheme, the symbols in Sat-DMIM become correlated due to the mode-selection mapping in the time domain. In this case, the *a priori* information of each Sat-DMIM symbol gleaned from the SISO decoder is subject to the mode-selection pattern constraint. As a further aggravation, if not fully compensated, the nonlinear dispersive satellite channels can significantly degrade the detection accuracy of both the mode-selection mapping symbols and the ordinary PSK/QAM constellation symbols. Therefore, the receiver has to jointly detect the IM bits and ordinary constellation mapping bits of LDPC-coded Sat-DMIM while coping with both linear and nonlinear ISI.

[d]Note that only odd-order terms appear in (3.1) , since the spectral components generated by even-order distortions fall outside of the channel bandwidth [92].

As illustrated in Fig. 3.1, by using the principle of Turbo equalizer [70], the SISO demodulator exploits the structured *a priori* information provided by the SISO decoder from the previous iteration to decouple the Sat-DMIM symbols. Given the estimates of Sat-DMIM symbols provided by the SISO demodulator, the demappers of the mode selector, as well as of the Mode A and Mode B, update the *a posteriori* log-likelihood ratios (LLRs) of coded bits within each subframe. By extracting the *a priori* information from the *a posteriori* probability distributions, the SISO decoder obtains the so-called *extrinsic* LLRs and then uses them to perform the standard Bahl, Cocke, Jelinek and Raviv (BCJR) decoding [73]. The output of the SISO decoder is then fed back to the SISO demodulator for the next iteration.

3.3 Amalgamated BP-MF-Based Iterative Detection Algorithm

3.3.1 Probabilistic Representation and Factor Graph

The goal of the optimal receiver is to find the estimates of information bits in $\mathbf{b}$ based on the maximum *a posteriori* (MAP) criterion:

$$\begin{aligned}\hat{b}_i^{\mathbf{MAP}} &= \arg\max_{b_i} p(b_i|\mathbf{r}) \\ &= \arg\max_{b_i} \sum\nolimits_{\mathbf{b}\setminus b_i} p(\mathbf{b}|\mathbf{r}), \quad i = 1, \ldots, M_i, \end{aligned} \tag{3.2}$$

where $\mathbf{r} \triangleq [r_1, \ldots, r_N]^T$ denotes the sequence of received signal samples. As can be easily analyzed, the criterion in (3.2) is not immune to the exponential complexity when a large block size M_i is involved in the presence of large channel memory. To avoid the intractable complexity, based on the received signal model in (3.1), we propose to factorize the joint *a posteriori* pdf of the information bits $\mathbf{b}$, the coded bits $\mathbf{c}$ and the Sat-DMIM symbols $\mathbf{x}$ as follows:

$$\begin{aligned} p(\mathbf{b}, \mathbf{c}, \mathbf{x}|\mathbf{r}) &\propto p(\mathbf{r}|\mathbf{x})p(\mathbf{x}|\mathbf{c})p(\mathbf{c}|\mathbf{b})p(\mathbf{b}) \\ &\propto \mathbb{I}_c(\mathbf{c}, \mathbf{b}) \prod_{g=1}^{G} f_{M,g}(\mathbf{x}_g|\mathbf{c}_g) P(\mathbf{x}) \prod_{n=1}^{N} f_{r,n}(x_n, x_{n-1}, \ldots, x_{n-L}, r_n), \end{aligned} \tag{3.3}$$

$$f_{r,n}(x_n,\ldots,x_{n-L},r_n) \propto \exp\left\{-\frac{1}{2\sigma^2}\Big|r_n-\sum_{l=0}^{L}x_{n-l}h_l-\sum_{i=0}^{L}\sum_{j\geq i}^{L}\sum_{k=0}^{L}x_{n-i}x_{n-j}x_{n-k}^{*}h_{ijk}\Big|^2\right\} \tag{3.4}$$

$$\begin{aligned}
&\propto \exp\Bigg\{-\frac{1}{2\sigma^2}\Bigg\{|x_n|^6|h_{000}|^2+|x_n|^4\Big[\psi_{n,n}^{o\lhd o,2}+\psi_{n,n}^{oo\lhd,2}+2\Re\left[h_{000}^{*}(h_0+\psi_{n,n}^{o\lhd\lhd})\right]\Big]+|x_n|^2\Big[|h_0|^2\\
&+\psi_{n,n}^{o\lhd\lhd,2}+\psi_{n,n}^{\lhd\lhd o,2}+2\Re\left[(-r_n+\varepsilon_{n,n})^{*}\psi_{n,n}^{o\lhd o}+h_0^{*}\psi_{n,n}^{o\lhd\lhd}\right]\Big]+|\varepsilon_{n,n}|^2+2\Re\Big[|x_n|^4x_n\left[h_{000}(\psi_{n,n}^{o\lhd o})^{*}\right.\\
&\left.+h_{000}^{*}\psi_{n,n}^{oo\lhd}\right]+|x_n|^2x_n^2\left[h_{000}(\psi_{n,n}^{\lhd\lhd o})^{*}+(\psi_{n,n}^{o\lhd o})^{*}\psi_{n,n}^{oo\lhd}\right]+x_n^2x_n^{*}[\psi_{n,n}^{oo\lhd}(h_0+\psi_{n,n}^{o\lhd\lhd})^{*}]\\
&+x_n^3\psi_{n,n}^{oo\lhd}(\psi_{n,n}^{\lhd\lhd o})^{*}+|x_n|^2x_n\left[(-r_n+\varepsilon_{n,n})^{*}h_{000}+(\psi_{n,n}^{o\lhd o})^{*}(h_0+\psi_{n,n}^{o\lhd\lhd})+\psi_{n,n}^{o\lhd o}(\psi_{n,n}^{\lhd\lhd o})^{*}\right]\\
&-r_n^{*}\varepsilon_{n,n}+x_n^2\left[(-r_n+\varepsilon_{n,n})^{*}\psi_{n,n}^{oo\lhd}+(h_0+\psi_{n,n}^{o\lhd\lhd})(\psi_{n,n}^{\lhd\lhd o})^{*}\right]\\
&+x_n\left[(-r_n+\varepsilon_{n,n})^{*}(h_0+\psi_{n,n}^{o\lhd\lhd})+(-r_n+\varepsilon_{n,n})(\psi_{n,n}^{\lhd\lhd o})^{*}\right]\Big]\Bigg\}\Bigg\}.
\end{aligned} \tag{3.5}$$

$$\begin{aligned}
&\psi_{n,n}^{o\lhd o}=\sum_{i=1}^{L}x_{n-i}h_{0i0},\qquad \psi_{n,n}^{oo\lhd}=\sum_{i=1}^{L}x_{n-i}^{*}h_{00i},\qquad \psi_{n,n}^{o\lhd\lhd}=\sum_{i=1}^{L}\sum_{j=1}^{L}x_{n-i}x_{n-j}^{*}h_{0ij},\\
&\psi_{n,n}^{o\lhd o,2}=\sum_{i=1}^{L}|x_{n-i}|^2|h_{0i0}|^2,\psi_{n,n}^{oo\lhd,2}=\sum_{i=1}^{L}|x_{n-i}^{*}|^2|h_{00i}|^2,\psi_{n,n}^{o\lhd\lhd,2}=\sum_{i=1}^{L}\sum_{j=1}^{L}|x_{n-i}|^2|x_{n-j}^{*}|^2|h_{0ij}|^2,\\
&\psi_{n,n}^{\lhd\lhd o}=\sum_{i=1}^{L}\sum_{j\geq i}^{L}x_{n-i}x_{n-j}h_{ij0},\psi_{n,n}^{\lhd\lhd o,2}=\sum_{i=1}^{L}\sum_{j\geq i}^{L}|x_{n-i}|^2|x_{n-j}|^2|h_{ij0}|^2,\\
&\varepsilon_{n,n}=\sum_{i=1}^{L}x_{n-i}h_i+\sum_{i=1}^{L}\sum_{j\geq i}^{L}\sum_{k=1}^{L}x_{n-i}x_{n-j}x_{n-k}^{*}h_{ijk}.
\end{aligned} \tag{3.6}$$

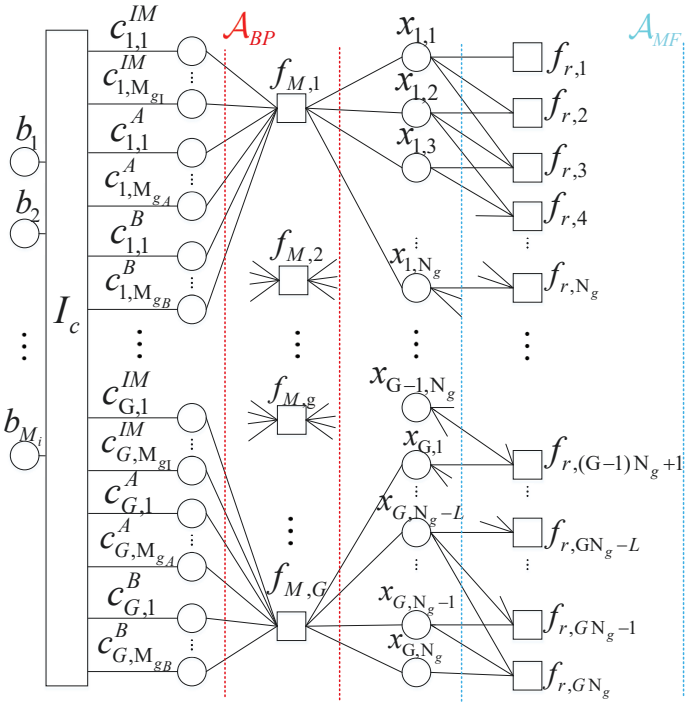

Figure 3.2: Factor graph representation of the probabilistic model (3.3).

where the indicator function $\mathbb{I}_c(\cdot)$ stands for the coding constraint and $f_{M,g}(\cdot)$ represents the mode-selection and PSK/QAM modulation constraints. In (3.3), the likelihood function $f_{r,n}(\cdot)$ is calculated by (3.5), where the variables $\psi^{o\lhd o}_{n,n}$, $\psi^{oo\lhd}_{n,n}$, $\psi^{o\lhd\lhd}_{n,n}$, $\psi^{\lhd\lhd o}_{n,n}$, $\psi^{o\lhd o,2}_{n,n}$, $\psi^{oo\lhd,2}_{n,n}$, $\psi^{o\lhd\lhd,2}_{n,n}$, $\psi^{\lhd\lhd o,2}_{n,n}$ and $\varepsilon_{n,n}$ that appear in (3.5) are given by (3.6). We observe that these variables are composed of Sat-DMIM symbols $\{x_{n-1},\ldots,x_{n-L}\}$ and are irrelated with x_n, as a result of classifying the linear and nonlinear ISI terms in (3.1) with reference to x_n. If an other variable x_m, $m=n-1,\ldots,n-L$, is the desired symbol, the function $f_{r,n}(\cdot)$ is supposed to be rewritten by rearranging the ISI terms, which will lead to an expression similar to (3.5). Note that the sufficient statistics of x_m, $m=n,\ldots,n-L$, involved in the likelihood function $f_{r_n}(\cdot)$ are determined by the order of Volterra model and stay the same for any dispersion L. For example, there are total $\mathcal{L}^2$ fixed sufficient statistics of x_m, including $|x_m|^6$, $|x_m|^4$, $|x_m|^2$, $\Re\{|x_m|^4x_m\}$, $\Re\{|x_m|^2x_m^2\}$, $\Re\{x_m^3\}$, $\Re\{|x_m|^2x_m\}$, $\Re\{x_m^2\}$ and $\Re\{x_m\}$, where $\mathcal{L}$ is the order of Volterra model. The factor graph representation corresponding to the factorization in (3.3) is shown in Fig. 3.2 for the case $L=2$.

As seen in Fig. 3.2, the Sat-DMIM symbols are correlated due to the mode-

selection pattern constraint and the memory in the overall satellite transponder response. The application of the standard sum-product algorithm [64] at both the constraint nodes $\{f_{M,g}\}_{g=1}^{G}$ and the observation nodes $\{f_{r,n}\}_{n=1}^{N}$ leads to the near-optimal forward-backward (FB) equalizer [88]. However, the implementation complexity of the FB equalizer is in the order of $\mathcal{O}((P_A + P_B)^L)$, which may be deemed excessive for large constellations in nonlinear channels with deep memory. To this end, we propose a low-complexity message passing algorithm by classifying the factors in Fig. 3.2 into two regions, i.e., the MF-region $\mathcal{A}_{\mathrm{MF}} = \{f_{r,n};\, n \in [1:N]\}$ and the BP-region $\mathcal{A}_{\mathrm{BP}} = \{f_{M,g};\, g \in [1:G]\}$, bearing in mind that the MF method is competitive with the exponential distribution family, whilst the BP algorithm delivers remarkable performance for discrete probabilistic models [67].

3.3.2 Message Passing at the Constraint Nodes

Instead of directly treating the correlated Sat-DMIM symbols as independent Gaussian variables, we explicitly construct the *a priori* information of Sat-DMIM symbols by taking full account of the mode-selection pattern constraint. Denote the *extrinsic* probabilities of coded bits provided by the SISO decoder as $\big\{P(c_{g,1}^{\mathrm{IM}}), \ldots, P(c_{g,M_{g_{\mathrm{I}}}}^{\mathrm{IM}}), P(c_{g,1}^{A}), \ldots, P(c_{g,M_{g_A}}^{A}), P(c_{g,1}^{B}), \ldots, P(c_{g,M_{g_B}}^{B})\big\}$. Then the *a priori* probabilities of the legitimate pattern and the ordinary PSK/QAM symbols are given by

$$\begin{aligned}
P(\mathcal{I}_g^A = \mathcal{I}^A(\gamma)) &= \prod_{i=1}^{M_{g_{\mathrm{I}}}} P(c_{g,i}^{\mathrm{IM}} = c_i^{\gamma}), \quad \gamma = 1, \ldots, \Upsilon, \\
P(x_g(\jmath) = \dot{s}_m) &= \prod_{i=1}^{\log_2 P_A} P(c_{g,(\jmath-1)\log_2 P_A + i}^{A} = c_i^{\dot{s}_m}), \quad \jmath \in \mathcal{I}^A(\gamma), m = 1, \ldots, P_A, \\
P(x_g(k) = \ddot{s}_l) &= \prod_{i=1}^{\log_2 P_B} P(c_{g,(k-1)\log_2 P_B + i}^{B} = c_i^{\ddot{s}_l}), \quad k \in \bar{\mathcal{I}}^A(\gamma), l = 1, \ldots, P_B,
\end{aligned} \tag{3.7}$$

where the sequences of coded bits $[c_1^{\gamma}, \ldots, c_{M_{g_{\mathrm{I}}}}^{\gamma}]^T$, $[c_1^{\dot{s}_m}, \ldots, c_{\log_2 P_A}^{\dot{s}_m}]^T$ and $[c_1^{\ddot{s}_l}, \ldots, c_{\log_2 P_B}^{\ddot{s}_l}]^T$ correspond to the γth legitimate pattern $\mathcal{I}^A(\gamma)$, the mth symbol $\dot{s}_m$ belonging to mode A and the lth symbol $\ddot{s}_l$ belonging to mode B, respectively. According to the process of mode-selection mapping as described in Section II, the *a priori*

$$D(x_i=\dot{s}_m) \propto \sum_{\gamma:i\in\mathcal{I}^A(\gamma)} P\big(\mathcal{I}_g^A=\mathcal{I}^A(\gamma)\big)\,\mu_{x_i\to f_{M,g}}(x_i)P(x_i=\dot{s}_m)$$
$$\cdot \prod_{\jmath\in\mathcal{I}^A(\gamma),\jmath\neq i}\left(\sum_{\dot{s}\in\dot{\mathcal{S}}}\mu_{x_\jmath\to f_{M,g}}(x_\jmath)P(x_\jmath=\dot{s})\right)\prod_{k\in\bar{\mathcal{I}}^A(\gamma)}\left(\sum_{\ddot{s}\in\ddot{\mathcal{S}}}\mu_{x_k\to f_{M,g}}(x_k)P(x_k=\ddot{s})\right), \quad \dot{s}_m\in\dot{\mathcal{S}},\quad m=1,\ldots,P_A. \tag{3.10}$$

$$E(x_i=\ddot{s}_l) \propto \sum_{\gamma:i\in\bar{\mathcal{I}}^A(\gamma)} P\left(\mathcal{I}_g^A=\bar{\mathcal{I}}^A(\gamma)\right)\mu_{x_i\to f_{M,g}}(x_i)P(x_i=\ddot{s}_l)$$
$$\cdot \prod_{\jmath\in\bar{\mathcal{I}}^A(\gamma),\jmath\neq i}\left(\sum_{\dot{s}\in\ddot{\mathcal{S}}}\mu_{x_\jmath\to f_{M,g}}(x_\jmath)P(x_\jmath=\ddot{s})\right)\prod_{k\in\mathcal{I}^A(\gamma)}\left(\sum_{\dot{s}\in\dot{\mathcal{S}}}\mu_{x_k\to f_{M,g}}(x_k)P(x_k=\dot{s})\right), \quad \ddot{s}_l\in\ddot{\mathcal{S}},\quad l=1,\ldots,P_B. \tag{3.11}$$

probabilities of the Sat-DMIM symbols can be constructed by

$$P(\mathbf{x})=\prod_{g=1}^{G}P_{\mathbf{x}_g}(\mathbf{x}_g),$$
$$P_{\mathbf{x}_g}(\mathbf{x}_g)=\sum_{\gamma=1}^{\Upsilon}P\big(\mathcal{I}_g^A=\mathcal{I}^A(\gamma)\big)\prod_{i\in\mathcal{I}^A(\gamma)}\left(\sum_{\dot{s}\in\dot{\mathcal{S}}}\delta(x_i-\dot{s})P(x_i)\right)\prod_{\jmath\in\bar{\mathcal{I}}^A(\gamma)}\left(\sum_{\ddot{s}\in\ddot{\mathcal{S}}}\delta(x_\jmath-\ddot{s})P(x_\jmath)\right), \tag{3.8}$$

where $i,\jmath\in I_g\triangleq\{(g-1)N_g+1:gN_g\}$ and $\bar{\mathcal{I}}^A(\gamma)$ is the complementary set of $\mathcal{I}^A(\gamma)$. Given (3.8) and the message $\mu_{x_i\to f_{M,g}}(x_i)$ which will be later updated according to (3.18), the *a posteriori* pdf $b(x_i)$ of the Sat-DMIM symbol x_i can be derived as

$$b(x_i=x)=\begin{cases}\frac{D(x_i=x)}{\sum_{\dot{s}\in\dot{\mathcal{S}}}D(x_i=\dot{s})+\sum_{\ddot{s}\in\ddot{\mathcal{S}}}E(x_i=\ddot{s})}, & x\in\dot{\mathcal{S}},\\ \frac{E(x_i=x)}{\sum_{\dot{s}\in\dot{\mathcal{S}}}D(x_i=\dot{s})+\sum_{\ddot{s}\in\ddot{\mathcal{S}}}E(x_i=\ddot{s})}, & x\in\ddot{\mathcal{S}},\end{cases} \tag{3.9}$$

where $D(\cdot)$ and $E(\cdot)$ are given by (3.10) and (3.11), respectively.

Then the *a posteriori* mean and variance of x_i are calculated by

$$\begin{aligned}\hat{x}_i&=\sum_{\dot{s}\in\dot{\mathcal{S}}}\dot{s}b(x_i=\dot{s})+\sum_{\ddot{s}\in\ddot{\mathcal{S}}}\ddot{s}b(x_i=\ddot{s}),\\ v_{x_i}&=\sum_{\dot{s}\in\dot{\mathcal{S}}}|\dot{s}|^2b(x_i=\dot{s})+\sum_{\ddot{s}\in\ddot{\mathcal{S}}}|\ddot{s}|^2b(x_i=\ddot{s})-|\hat{x}_i|^2.\end{aligned} \tag{3.12}$$

Note that for the first iteration, there is no *extrinsic* information arriving from the SISO decoder. By designing the legitimate pattern set properly, the structured *a*

priori pdf of $\mathbf{x}_g$ can be directly written as

$$P^0_{\mathbf{x}_g}(\mathbf{x}_g)=\frac{1}{\Upsilon}\sum_{\gamma=1}^{\Upsilon}\left[\prod_{i\in\mathcal{I}^A(\gamma)}\left(\frac{1}{P_A}\sum_{\dot{s}\in\dot{\mathcal{S}}}\delta(x_i-\dot{s})\right)\cdot\prod_{\jmath\in\bar{\mathcal{I}}^A(\gamma)}\left(\frac{1}{P_B}\sum_{\ddot{s}\in\ddot{\mathcal{S}}}\delta(x_\jmath-\ddot{s})\right)\right] \tag{3.13}$$

Accordingly, the initial *a priori* mean and variance of x_i are given by

$$\begin{aligned}\hat{x}_i^0&=\sum_{\dot{s}\in\dot{\mathcal{S}}}\dot{s}\frac{K_g}{P_AN_g}+\sum_{\ddot{s}\in\ddot{\mathcal{S}}}\ddot{s}\frac{N_g-K_g}{P_BN_g}\\ v_{x_i}^0&=\sum_{\dot{s}\in\dot{\mathcal{S}}}|\dot{s}|^2\frac{K_g}{P_AN_g}+\sum_{\ddot{s}\in\ddot{\mathcal{S}}}|\ddot{s}|^2\frac{N_g-K_g}{P_BN_g}-|\hat{x}_i^0|^2.\end{aligned} \tag{3.14}$$

3.3.3 Message Passing at the Observation Nodes

According to the MF rule [67], we can calculate the message $\mu_{f_{r,m}\to x_n}(x_n)$ forwarded from the observation node $f_{r,m}$ to the Sat-DMIM symbol variable node x_n, $n\in[m,m-L]$, by

$$\mu_{f_{r,m}\to x_n}(x_n)\propto\exp\Big\{E_{\prod_{m\le i\le m-L,i\neq n}b(x_i)}\log f_{r,m}(x_m,\ldots,x_{m-L},r_m)\Big\} \tag{3.15}$$

$$\begin{aligned}&\propto\exp\Big\{-\frac{1}{2\sigma^2}\big(\hat{\varphi}^{3,3}_{r_m,n}|x_n|^6+\hat{\varphi}^{2,2}_{r_m,n}|x_n|^4+\hat{\varphi}^{1,1}_{r_m,n}|x_n|^2+2\Re\{\hat{\phi}^{3,2}_{r_m,n}|x_n|^4x_n\\ &\qquad+\hat{\phi}^{3,1}_{r_m,n}|x_n|^2x_n^2+\hat{\phi}^{3,0}_{r_m,n}x_n^3+\hat{\phi}^{2,1}_{r_m,n}|x_n|^2x_n+\hat{\phi}^{2,0}_{r_m,n}x_n^2+\hat{\phi}^{1,0}_{r_m,n}x_n\}\big)\Big\}\\ &\triangleq\exp\Big\{-\frac{1}{2\sigma^2}\Big(\sum_{q,v}\hat{\varphi}^{q,v}_{r_m,n}x_n^q(x_n^*)^v+2\Re\Big\{\sum_{k,l}\hat{\phi}^{k,l}_{r_m,n}x_n^k(x_n^*)^l\Big\}\Big)\Big\},\end{aligned} \tag{3.16}$$

where the *a posteriori* pdf $b(x_i)$ of x_i is updated in the previous iteration using (4.49); the parameters $\hat{\varphi}^{q,v}_{r_m,n}$ and $\hat{\phi}^{k,l}_{r_m,n}$ consist of the products of the Volterra kernels and the first- and second-order moments of variables $\{x_m,\ldots,x_{m-L}\}$ except x_n; and the parameters $\{q,v\}$ and $\{k,l\}$ indicate the number of variables x_n and x_n^* involved in the sufficient statistics that associated with the parameters $\hat{\varphi}^{q,v}_{r_m,n}$ and $\hat{\phi}^{k,l}_{r_m,n}$. Considering that the Sat-DMIM symbols are discrete variables, the parameters $\hat{\varphi}^{q,v}_{r_m,n}$ and $\hat{\phi}^{k,l}_{r_m,n}$ can be calculated by using the enumeration method [100]. For example, when $m=n$, the parameter associated with $|x_n|^4$, i.e., $\hat{\varphi}^{2,2}_{r_n,n}=\hat{\psi}_n^{o\triangleleft o,2}+\hat{\psi}_n^{oo\triangleleft,2}+2\Re[h^*_{000}(h_0+\hat{\psi}_n^{o\triangleleft\triangleleft})]$, with $\hat{\psi}_n^{o\triangleleft o,2}\triangleq\sum_{i=1}^L(|\hat{x}_{n-i}|^2+v_{x_{n-i}})|h_{0i0}|^2$, $\hat{\psi}_n^{oo\triangleleft,2}\triangleq\sum_{i=1}^L(|\hat{x}_{n-i}|^2+v_{x_{n-i}})|h_{00i}|^2$ and $\hat{\psi}_n^{o\triangleleft\triangleleft}\triangleq\sum_{i=1}^L\sum_{j=1}^L\hat{x}_{n-i}\hat{x}^*_{n-j}h_{0ij}$; and the paramter associated with $2\Re\{|x_n|^4x_n\}$, i.e., $\hat{\phi}^{3,2}_{r_n,n}=h_{000}(\hat{\psi}_n^{o\triangleleft o})^*+h^*_{000}\hat{\psi}_n^{oo\triangleleft}$, with $\hat{\psi}_n^{o\triangleleft o}\triangleq\sum_{i=1}^L\hat{x}_{n-i}h_{0i0}$ and $\hat{\psi}_n^{oo\triangleleft}\triangleq\sum_{i=1}^L\hat{x}^*_{n-i}h_{00i}$.

Given the message $\mu_{f_{r,m}\to x_n}(x_n)$, $m = n, \ldots, n+L$, the message forwarded from the variable node x_n to the constraint node $f_{M,g}$ can then be updated by

$$\mu_{x_n\to f_{M,g}}(x_n) \propto \prod_{m=n}^{n+L} m_{f_{r,m}\to x_n}(x_n) \tag{3.17}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{m=n}^{n+L}\left(\sum_{q,v}\hat{\varphi}_{r_m,n}^{q,v}x_n^q(x_n^*)^v + 2\Re\Big\{\sum_{k,l}\hat{\phi}_{r_m,n}^{k,l}x_n^k(x_n^*)^l\Big\}\right)\right\}, \tag{3.18}$$

where $n \in I_g$. It is readily seen that updating the messages $\mu_{f_{r,m}\to x_n}(x_n)$, $m = n, \ldots, n+L$, and $\mu_{x_n\to f_{M,g}}(x_n)$, $n = 1, \ldots, N$, only requires updating the parameters of them by calculating the products of specific combinations of the first- and second-order moments of L consecutive variables within each Sat-DMIM frame.

By substituting the message $\mu_{x_n\to f_{M,g}}(x_n)$ into (3.10) and (3.11), the belief of x_n can be immediately updated according to (4.49). Meanwhile, we obtain the belief of $\mathbf{x}_g$ by

$$b(\mathbf{x}_g) = P_{\mathbf{x}_g}(\mathbf{x}_g)\prod_{i\in I_g}\mu_{x_i\to f_{M,g}}(x_i). \tag{3.19}$$

Then the *extrinsic* LLRs of the ith mode-selection mapping bit $c_{g,i}^{\mathrm{IM}}$, the jth mode A mapping bit $c_{g,j}^{A}$ and the kth mode B mapping bit $c_{g,k}^{B}$ associated with the gth Sat-DMIM symbol vector $\mathbf{x}_g$, i.e., $L^{\mathrm{extr}}(c_{g,i}^{\mathrm{IM}})$, $L^{\mathrm{extr}}(c_{g,j}^{A})$ and $L^{\mathrm{extr}}(c_{g,k}^{B})$, respectively, can be iteratively updated based on the Turbo principle [70].

3.3.4 Message Passing Scheduling

The message passing schedule for the proposed joint Sat-DMIM symbols detection and decoding algorithm based on the amalgamated BP-MF method in nonlinear dispersive satellite channels, which is referred to as BP-MF, is summarized in Algorithm 5.

Algorithm 5 The Proposed Joint Sat-DMIM Symbols Detection and Decoding Algorithm Based on the Amalgamated BP-MF Method in Nonlinear Dispersive Satellite Channels.

1: Initialization: $t = 0$ (iteration index)

2: For $n = 1, \ldots, N$, compute $\hat{x}_n^0$ and $v_{x_n}^0$ according to (3.14).

3: Set $t = 1$ and repeat the following steps until $\sum_n |\hat{x}_n^{t+1} - \hat{x}_n^t|^2 < \epsilon \sum_n |\hat{x}_n^t|^2$ or $t > T_{\max}$, where ϵ is a pre-specified error tolerance and $T_{\max}$ is the maximum number of iterations.

4: − Decoupling step: For $n = 1, \ldots, N$, $m = n, \ldots, n + L$, compute the message $\mu_{f_{r,m} \to x_n}(x_n)$ and $\mu_{x_n \to f_{M,g}}(x_n)$ from (3.16) and (3.18), respectively.

5: − Decoding step: For $g = 1 \ldots, G$, $i = 1, \ldots, M_{g_{\mathrm{I}}}$, $j = 1, \ldots, M_{g_A}$ and $k = 1, \ldots, M_{g_B}$, compute the *extrinsic* LLRs $L^{\mathrm{extr}}(c_{g,i}^{\mathrm{IM}})$, $L^{\mathrm{extr}}(c_{g,j}^{A})$ and $L^{\mathrm{extr}}(c_{g,k}^{B})$, and then forward them to the SISO decoder to perform the standard BCJR decoding [73].

6: Compute (3.7) and (3.8) by using the *extrinsic* information emerging from the SISO decoder.

7: − Denoising step: For $n = 1, \ldots, N$, compute $\{\hat{x}_n^t, v_{x_n}^t\}$ from (4.50).

3.4 Partial Linearization Approximation-Based Iterative Detection Algorithm

3.4.1 Complexity Reduction Using the Simplification of Message Parameters

Observe by referring to (3.6), (3.16) and (3.18), that $\mathcal{O}(L+1)(\frac{1}{2}L^3+\frac{7}{2}L^2+6L)$ multiplications are involved for detecting each Sat-DMIM symbol. This means that the computational complexity of the amalgamated BP-MF algorithm is still considerable for large channel memory, since it grows with the fourth power of the channel dispersion. Therefore, a reduced-complexity iterative detection algorithm is required for simplifying the updating process of messages $\mu_{f_{r,m}\to x_n}(x_n)$ and $\mu_{x_n\to f_{M,g}}(x_n)$ from (3.16) and (3.18), respectively. Due to the very similar forms of messages $\{\mu_{f_{r,m}\to x_n}(x_n)\}_{m=n}^{n+L}$, in the following, the approximation of the message $\mu_{f_{r,n}\to x_n}(x_n)$ is presented, and similar results hold for other messages. Considering that these messages all have compact parametric representations, the proposed low-complexity message passing method reduces to approximating and propagating the function parameters. Firstly, the estimate of variable $\varepsilon_{n,n}$ in (3.6) is approximated as

$$\hat{\varepsilon}_{n,n} \approx \hat{\underline{\varepsilon}}_{n,n} \triangleq \sum_{i=1}^{L} \hat{x}_{n-i} h_i, \tag{3.20}$$

assuming $h_{ijk}=0$ for $i,\jmath,k\neq 0$. This operation is equivalent to discarding some high-order terms in the message $\mu_{f_{r,n}\to x_n}(x_n)$. Note that the approximation in (3.20) is different from the nonlinear soft interference canceler of [86] which directly neglects the contribution of third-order Volterra terms. In the proposed approximation method, not all h_{ijk}, $i,\jmath,k\neq 0$, are treated as zeros when calculating (3.18), because these Volterra kernels are involved in calculating messages $\{\mu_{f_{r,m}\to x_n}(x_n)\}_{m=n+1}^{n+L}$ from (3.16) and therefore contribute to the update of messages $\{\mu_{x_n\to f_{M,g}}(x_n)\}_{n=1}^{N}$ from (3.18). For example, the Volterra kernels h_{ijk}, $i,\jmath,k\neq 0$ are used to update the message $\mu_{f_{r,n+1}\to x_n}(x_n)$, as long as neither i, $\jmath$ nor k equals to 1. By resorting to this approximation, $\mathcal{O}(L+1)(\frac{1}{2}L^2(L+1))$ operations of calculating the products of

Volterra kernels and the first-order moments of Sat-DMIM symbols are omitted in the calculations of (3.18).

Secondly, the estimates of $\psi_{n,n}^{o\lhd\lhd,2}$ and $\psi_{n,n}^{\lhd\lhd o,2}$ are approximated as

$$\begin{aligned}\hat{\psi}_{n,n}^{o\lhd\lhd,2} &\approx \underline{\psi}_{n,n}^{o\lhd\lhd,2} \triangleq e^2 \sum_{i=1}^{L}\sum_{\jmath=1}^{L} |h_{0i\jmath}|^2,\\ \hat{\psi}_{n,n}^{\lhd\lhd o,2} &\approx \underline{\psi}_{n,n}^{\lhd\lhd o,2} \triangleq e^2 \sum_{i=1}^{L}\sum_{\jmath\geq i}^{L} |h_{i\jmath 0}|^2,\end{aligned} \tag{3.21}$$

where e is the average energy per transmitted Sat-DMIM symbol. In this way, another $\mathcal{O}(L+1)(L^2+\frac{1}{2}L(L+1))$ operations are further omitted. Note that this is different from the direct approximation of the system model in [87], where $h_{0i\jmath},\ i,\jmath = 1,\ldots,L$, and $h_{i\jmath 0},\ i,\jmath = 1,\ldots,L,\ \jmath \geq i$, are directly replaced by zeros. As seen in (3.6), $h_{0i\jmath}$ and $h_{i\jmath 0}$ still contribute to the update of parameters $\psi_{n,n}^{o\lhd\lhd}$ and $\psi_{n,n}^{\lhd\lhd o}$. Moreover, in our proposed approximation method, the second-order moments of Sat-DMIM symbols are updated in an iterative fashion, i.e., $E_{b(x_{n-i})}[|x_{n-i}|^2] = |\hat{x}_{n-i}|^2 + v_{x_{n-i}},\ i = 1,\ldots,L$. These values are then used to calculate the parameters $\hat{\psi}_{n,n}^{o\lhd o,2}$ and $\hat{\psi}_{n,n}^{oo\lhd,2}$ of the message $\mu_{f_{r,n}\to x_n}(x_n)$. By contrast, the second-order moments of transmitted symbols except of the desired one are approximated by the average energy and are not updated with the iterations in [87]. On the other hand, considering that the *a posteriori* variances of variables $\{x_{n-i}\}_{i=1}^{L}$, i.e., $\{v_{x_{n-1}}\}_{i=1}^{L}$, are explicitly exploited to update the message $\mu_{f_{r,n}\to x_n}(x_n)$ through the parameters $\hat{\psi}_{n,n}^{o\lhd o,2}$ and $\hat{\psi}_{n,n}^{oo\lhd,2}$, the approximation in (3.21) is different from directly approximating the nonlinear terms as Gaussian noise and then using $\{v_{x_{n-1}}\}_{i=1}^{L}$ merely to compute the equivalent variance of the Gaussian noise process in [93]. From (3.20) and (3.21), it can be found that the number of multiplications required for tracking the message $\mu_{x_n\to f_{M,g}}(x_n)$ is reduced from $\mathcal{O}(\frac{1}{2}L^4+4L^3+\frac{19}{2}L^2+6L)$ to $\mathcal{O}(\frac{3}{2}L^3+7L^2+\frac{11}{2}L)$ per Sat-DMIM symbol.

3.4.2 Complexity Reduction Using the First-Order Taylor Series Expansion Approximation

The message $\mu_{x_n\to f_{M,g}}(x_n)$ in (3.18) is a nonlinear distribution of variable x_n and consists of $\mathcal{L}^2$ sufficient statistics of x_n. By substituting (3.18) into (3.10) and (3.11),

$$\mu_{f_{r,n}\to x_n}(x_n) \propto \exp\Big\{E_{b(x_{n-1}),\ldots,b(x_{n-L})}\Big[-\frac{1}{2\sigma^2}\big|r_n-\big(f_n(x_n,x_n^*)+x_n(h_0+\psi_{n,n}^{o\lhd\lhd})+x_n^*\psi_{n,n}^{\lhd\lhd o}+\varepsilon_{n,n}\big)\big|^2\Big]\Big\}$$

$$\approx \exp\Big\{E_{b(x_{n-1}),\ldots,b(x_{n-L})}\Big[-\frac{1}{2\sigma^2}\big|r_n-\big(x_n[2\hat{x}_n\hat{x}_n^*h_{000}+2\hat{x}_n\psi_n^{oo\lhd}+\hat{x}_n^*\psi_{n,n}^{o\lhd o}+h_0+\psi_{n,n}^{o\lhd\lhd}]+x_n^*[\hat{x}_n^2h_{000}+\hat{x}_n\psi_{n,n}^{o\lhd o}+\psi_{n,n}^{\lhd\lhd o}]-2\hat{x}_n^2\hat{x}_n^*h_{000}-\hat{x}_n^2\psi_{n,n}^{oo\lhd}-\hat{x}_n\hat{x}_n^*\psi_{n,n}^{o\lhd o}+\varepsilon_{n,n}\big)\big|^2\Big]\Big\}$$

$$\approx \exp\Big\{-\frac{1}{2\sigma^2}\Big[|x_n|^2\big[5|\hat{x}_n|^4|h_{000}|^2+|\hat{x}_n|^2(2\hat{\psi}_{n,n}^{o\lhd o,2}+4\hat{\psi}_{n,n}^{oo\lhd,2})+|h_0|^2+\underline{\psi}_{n,n}^{o\lhd\lhd,2}+\underline{\psi}_{n,n}^{\lhd\lhd o,2}$$
$$+2\Re\{|\hat{x}_n|^2\hat{x}_n[4h_{000}^*\hat{\psi}_{n,n}^{oo\lhd}+3h_{000}\hat{\psi}_{n,n}^{o\lhd o,2}]+2\hat{x}_n^2\hat{\psi}_{n,n}^{oo\lhd}(\hat{\psi}_{n,n}^{o\lhd o})^*+2|\hat{x}_n|^2h_{000}^*(h_0+\underline{\psi}_{n,n}^{o\lhd\lhd})$$
$$+\hat{x}_n[2\hat{\psi}_{n,n}^{oo\lhd}(h_0+\underline{\psi}_{n,n}^{o\lhd\lhd})^*+(\hat{\psi}_{n,n}^{o\lhd o})^*(h_0+\underline{\psi}_{n,n}^{o\lhd\lhd})+h_{000}(\underline{\psi}_{n,n}^{\lhd\lhd o})^*+\hat{\psi}_{n,n}^{o\lhd o}(\underline{\psi}_{n,n}^{\lhd\lhd o})^*]+h_0^*\underline{\psi}_{n,n}^{o\lhd\lhd}\}\big]$$
$$+2\Re\big\{x_n^2[\hat{\xi}_{n,n}\hat{\zeta}_{n,n}^*]+x_n[\hat{\xi}_{n,n}(\hat{\rho}_{n,n}+\hat{\underline{\varepsilon}}_{n,n})^*+\hat{\zeta}_{n,n}^*(\hat{\rho}_{n,n}+\hat{\underline{\varepsilon}}_{n,n})-r_n^*\hat{\xi}_{n,n}-r_n\hat{\zeta}_{n,n}^*]\big\}\Big]\Big\}$$
$$\propto \exp\Big\{-\frac{1}{2\sigma^2}\Big[\hat{\underline{\varphi}}_{r_n,n}^{1,1}|x_n|^2+2\Re\big\{\hat{\underline{\phi}}_{r_n,n}^{2,0}x_n^2+\hat{\underline{\phi}}_{r_n,n}^{1,0}x_n\big\}\Big]\Big\}. \tag{3.22}$$

it is seen that the complexity of calculating the belief of Sat-DMIM symbol from (4.49) is $\mathcal{O}(\mathcal{L}^2(P_A+P_B))$. By contrast, the complexity in this part of classical SC signal transmitting over linear dispersive channels only grows linearly with the modulation order. To reduce the complexity of this part in Sat-DMIM, the first-order Taylor series expansion is used to further approximate the messages $\{\mu_{f_{r,m}\to x_n}(x_n)\}_{m=n}^{n+L}$. Without loss of generality, the derivations of the approximate message $\mu_{f_{r,n}\to x_n}(x_n)$ are provided, and other messages can be updated with a similar idea. Firstly, by regarding x_n as the desired variable, the nonlinear terms associated with x_n in the log-likelihood $\log f_{r,n}(x_n,\ldots,x_{n-L},r_n)$ can be written as

$$\begin{aligned} f_n(x_n,x_n^*) &= |x_n|^2x_nh_{000}+|x_n|^2\big(\sum\nolimits_{i=1}^{L}x_{n-i}h_{0i0}\big)+x_n^2\big(\sum\nolimits_{i=1}^{L}x_{n-i}^*h_{00i}\big)\\ &= x_nx_nx_n^*h_{000}+x_nx_n\psi_{n,n}^{oo\lhd}+x_nx_n^*\psi_{n,n}^{o\lhd o}, \end{aligned} \tag{3.23}$$

which is subject to the constraints $\Re\{x_n\}=\Re\{x_n^*\}$ and $\Im\{x_n\}+\Im\{x_n^*\}=0$. Here, in contrast to denoting $x_nx_n^*$ by $|x_n|^2$ as in (3.5), the sufficient statistics of x_n is rewritten in an expanded form. This apparently fine distinction is of great necessity to obtain an analytic expression without encountering meaningless expressions, such as $\partial(|x_n|^2x_n)/\partial x_n$. To elaborate a little further, assuming a function $f:\mathbb{C}\to\mathbb{C}$ with x^* as its variable, for example, $f(x)=x^*$, it is not analytic, i.e., the differentiation of x^* by x is not defined in normal complex-variable theory [101]. Thus the real function

$f : \mathbb{C} \to \mathbb{R}$ given by $f(x) = |x|^2$ is not analytic either. Fortunately, the function $f_n(x_n, x_n^*)$ in (3.23) is analytic with respect to x_n and x_n^* independently in the sense of partial differentiation. However, calculating the complex-variable derivative is quite a stringent task that is subject to the Cauchy-Riemann conditions [102]. To solve this problem, the following theorem is introduced, the detailed proof of which is provided in [101].

Theorem 1: Let $f : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ be a function of a complex variable x and its conjugate x^*, where $x = y + jz$, y and z are real numbers and $j = \sqrt{-1}$. Then the partial complex-variable derivatives $\partial f/\partial x$ (treating x^* as a constant in f) and $\partial f/\partial x^*$ (treating x as a constant in f) are equivalent to $\frac{1}{2}(\partial f/\partial y - j\partial f/\partial z)$ and $\frac{1}{2}(\partial f/\partial y + j\partial f/\partial z)$, respectively.

Given above theorem, the cumbersome expressions of complex-variable derivatives derived in the real field can be mathematically securely denoted by compact complex-field expressions for the sake of convenience. Thus, the first-order Taylor series expansion of $f_n(x_n, x_n^*)$ at the estimate of x_n gives

$$\delta f_n(x_n, x_n^*) \approx \frac{\partial f_n(x_n, x_n^*)}{\partial x_n}\delta x_n + \frac{\partial f_n(x_n, x_n^*)}{\partial x_n^*}\delta x_n^*, \tag{3.24}$$

where the partial derivative $\frac{\partial f_n(x_n, x_n^*)}{\partial x_n}$ (alternatively, $\frac{\partial f_n(x_n, x_n^*)}{\partial x_n^*}$) is obtained by treating x_n^* (alternatively, x_n) as a constant in f_n. By substituting (3.24) into the function $\log f_{r,n}(x_n, \ldots, x_{n-L}, r_n)$ in (3.15) and using the approximations in (3.20) and (3.21), we arrive at the partial-linearization approximation of the message $\mu_{f_{r,n}\to x_n}(x_n)$ in (3.22), where the parameters $\hat{\xi}_{n,n}$, $\hat{\zeta}_{n,n}$, and $\hat{\rho}_{n,n}$ are made up of the estimate of x_n obtained from the previous iteration, i.e., $\hat{\xi}_{n,n} = 2|\hat{x}_n|^2 h_{000} + 2\hat{x}_n\hat{\psi}_{n,n}^{oo\triangleleft} + \hat{x}_n^*\hat{\psi}_{n,n}^{o\triangleleft o} + h_0 + \underline{\psi}_{n,n}^{o\triangleleft\triangleleft}$, $\hat{\zeta}_{n,n} = \hat{x}_n^2 h_{000} + \hat{x}_n\hat{\psi}_{n,n}^{o\triangleleft o} + \underline{\psi}_{n,n}^{\triangleleft\triangleleft o}$, and $\hat{\rho}_{n,n} = -2|\hat{x}_n|^2\hat{x}_n h_{000} - \hat{x}_n^2\hat{\psi}_{n,n}^{oo\triangleleft} - |\hat{x}_n|^2\hat{\psi}_{n,n}^{o\triangleleft o}$.

In a similar way, the number of sufficient statistics of x_n involved in messages $\{\mu_{f_{r,m}\to x_n}(x_n)\}_{n+1}^{n+L}$ can also be reduced. After some algebraic manipulations, we find that the preserved sufficient statistics of these messages are the same, which are $|x_n|^2$, $\Re\{x_n^2\}$ and $\Re\{x_n\}$. Denote the parameters of the message $\mu_{f_{r,m}\to x_n}(x_n)$ corresponding to $|x_n|^2$, $\Re\{x_n^2\}$ and $\Re\{x_n\}$ by $\hat{\underline{\varphi}}_{r_m,n}^{1,1}$, $\hat{\underline{\phi}}_{r_m,n}^{2,0}$ and $\hat{\underline{\phi}}_{r_m,n}^{1,0}$, respectively. Then the message $\mu_{x_n\to f_{M,g}}(x_n)$ can be approximately updated by

$$\mu_{x_n\to f_{M,g}}(x_n) \approx \exp\left\{-\frac{1}{2\sigma^2}\sum_{m=n}^{n+L}\left(\hat{\underline{\varphi}}_{r_m,n}^{1,1}|x_n|^2 + 2\Re\left\{\hat{\underline{\phi}}_{r_m,n}^{2,0}x_n^2 + \hat{\underline{\phi}}_{r_m,n}^{1,0}x_n\right\}\right)\right\}. \tag{3.25}$$

Table 3.2: Comparisons of computational complexity (For example: $L = 3$, $P_A = 4$, $P_B = 8$ and $\beta = 4$.)

Algorithm	Complexity of the demodulator	Value
BP-MF	$\mathcal{O}((L+1)(\frac{1}{2}L^3 + \frac{7}{2}L^2 + 6L) + \mathcal{L}^2(P_A + P_B))$	$\mathcal{O}(360)$
PL-BP-MF	$\mathcal{O}(\frac{3}{2}L^3 + 7L^2 + \frac{11}{2}L + 3(P_A + P_B))$	$\mathcal{O}(156)$
FB [88]	$\mathcal{O}((P_A + P_B)^L)$	$\mathcal{O}(1.7 \times 10^3)$
Linear MMSE [70]	$\mathcal{O}(\beta(\beta^2 + P_A + P_B))$	$\mathcal{O}(112)$
Nonlinear MMSE [94]	$\mathcal{O}(\beta(\beta^2 + (\frac{1}{2}L^3 + \frac{7}{2}L^2 + 6L) + (P_A + P_B)))$	$\mathcal{O}(364)$

Using (3.22) and (3.25), the complexity order of the decoupling step in Algorithm 5 is reduced to $\mathcal{O}(3(P_A + P_B))$.

3.4.3 Message Passing Scheduling

The message passing schedule for the proposed joint Sat-DMIM symbols detection and decoding algorithm based on the partial-linearization approximation-aided amalgamated BP-MF method in nonlinear dispersive satellite channels, which is referred to as PL-BP-MF, is summarized in Algorithm 6.

Algorithm 6 The Proposed Joint Sat-DMIM Symbols Detection and Decoding Algorithm Based on the Partial-Linearization Approximation-Aided Amalgamated BP-MF Method in Nonlinear Dispersive Channels.

1: The same as BP-MF except for the following modifications imposed on the the decoupling step:

2: − Decoupling step: For $n = 1, \ldots, N$, $m = n, \ldots, n + L$, compute the message $\mu_{f_{r,m} \to x_n}(x_n)$ and $\mu_{x_n \to f_{M,g}}(x_n)$ from (3.22) and (3.25), respectively.

3.4.4 Complexity Analysis

In Table 3.2, the complexity of the proposed BP-MF and PL-BP-MF methods is briefly compared with existing equalizers in nonlinear dispersive satellite channels. In code-aided systems, all of these methods need to perform the standard BCJR decoding [73]. Thus only the computational complexity in the soft demodulation part needs to be considered. The near-optimal FB equalizer of [88] extended from the conventional Satcom to the proposed Sat-DMIM scheme has a complexity order

of $\mathcal{O}((P_A + P_B)^L)$. By ignoring the nonlinear ISI terms, the conventional linear equalizer of [70] requires $\mathcal{O}(\beta(\beta^2 + P_A + P_B))$ operations, where β is the size of the sliding window. In [94], the BER performance of the linear equalizer is further improved by taking the nonlinear ISI terms into account, at the cost of increasing the complexity order to $\mathcal{O}(\beta(\beta^2 + (\frac{1}{2}L^3 + \frac{7}{2}L^2 + 6L) + (P_A + P_B)))$. For the proposed amalgamated BP-MF method, $\mathcal{O}((L+1)(\frac{1}{2}L^3 + \frac{7}{2}L^2 + 6L))$ operations of calculating the products of the Volterra kernels and the first/second-order moments of Sat-DMIM symbols are required for evaluating (3.16), and $\mathcal{O}(\mathcal{L}^2(P_A + P_B))$ operations are needed for calculating (3.10) and (3.11). With the aid of the partial-linearization approximations in (3.20), (3.21) and (3.24), the total number of multiplications involved in the proposed PL-BP-MF method are reduced to $\mathcal{O}(\frac{3}{2}L^3 + 7L^2 + \frac{11}{2}L)$, and accordingly the complexity order of calculating (3.10) and (3.11) is reduced to $\mathcal{O}(3(P_A + P_B))$. In Table II, an example is provided for $L = 3$, $P_A = 4$, $P_B = 8$ and $\beta = 4$. It is observed that the proposed BP-MF and PL-BP-MF methods are computationally more efficient compared to the FB equalizer. The complexity order of BP-MF is slightly lower than that of the existing nonlinear equalizer.

3.5 Simulation Results

In this section, the performance of the proposed iterative detection algorithms for the Sat-DMIM scheme is evaluated by Monte Carlo simulations. In all simulations, a rate-2/3 (i.e., $R_c = 2/3$) LDPC code[e] with a block size 6048 is employed. The maximum number of iterations within the LDPC decoder is 50, while it is $T_{\max} = 20$ iterations between the SISO demodulation and SISO decoding, unless otherwise stated. The tolerance value is set to be $\epsilon = 10^{-12}$. For characterizing the nonlinear dispersive satellite channels, a third-order Volterra model (denoted by "Nonlinear Channel-I") is considered, the coefficients of which were derived in [92] but with stronger nonlinear ISI [94]. The average channel energy is normalized to 1.

Fig. 3.3 shows the BERs of the proposed BP-MF algorithm for the Sat-DMIM

[e]The variable and check node degree distributions are $X_1 = 0.00005 + 0.2043X + 0.5037X^2 + 0.2902X^7$ and $X_2 = 0.1975X^8 + 0.8025X^9$, respectively. The parameter definitions follow [72].

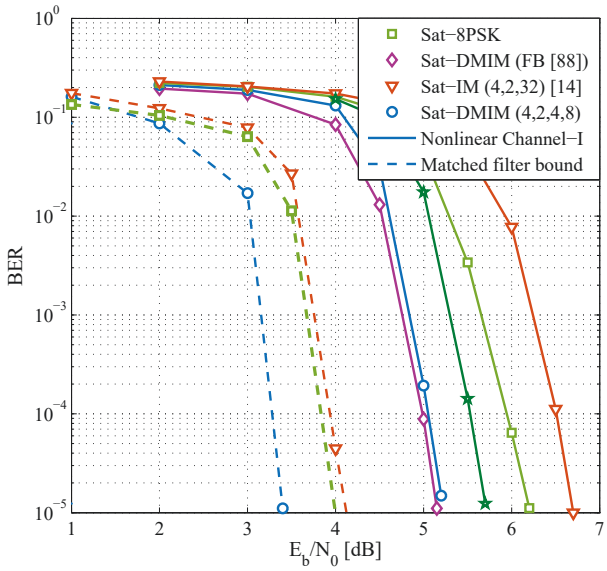

Figure 3.3: BER performance of classical 8PSK modulated Satcom, Sat-DMIM (characterized by "$(4, 2, 4, 8)$") and Sat-IM (characterized by "$(4, 2, 32)$") systems communicating over nonlinear dispersive satellite channels modeled by Volterra model "Nonlinear Channel-I" [94]. Both the proposed BP-MF and the existing FB [88] methods are evaluated. The effective throughput is 2 bits/symbol.

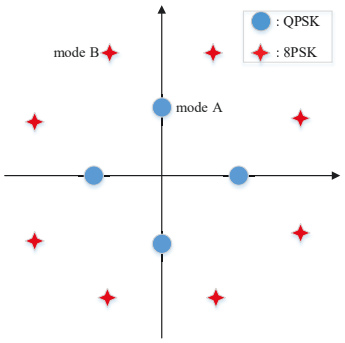

Figure 3.4: The constellation design of the Sat-DMIM scheme (characterized by "$(4, 2, 4, 8)$") following [44].

scheme characterized by $N_g = 4$, $K_g = 2$, $P_A = 4$ and $P_B = 8$ (denoted by "Sat-DMIM $(4, 2, 4, 8)$"), i.e., the modes A and B are QPSK and 8PSK constellations, respectively. The Volterra model Nonlinear Channel-I having a channel memory of $L = 3$ symbols is considered [94]. In order to increase the minimum Euclidean distance of the joint constellation set $\dot{\mathcal{S}} \cup \ddot{\mathcal{S}}$, the 8PSK constellation is rotated by a phase of $\pi/8$ [44]. The constellation design of the Sat-DMIM $(4, 2, 4, 8)$ scheme is shown in Fig. 3.4, where the average transmit power of the joint constellation is fixed to unity, i.e., $\dot{\mathcal{S}} = \{0.5671 + j0, 0 + j0.5671, -0.5671 + j0, 0 - j0.5671\}$ and $\ddot{\mathcal{S}} = \{1.1974 + j0.4960, 0.4960 + j1.1974, -0.4960 + j1.1974, -1.1974 + j0.4960, -1.1974 - j0.4960, -0.4960 - j1.1974, 0.4960 - j1.1974, 1.1974 - j0.4960\}$. For comparison, the BER curve of conventional 8PSK modulated Satcom (denoted by "Sat-8PSK"), yielding the same transmission rate as that of Sat-DMIM $(4, 2, 4, 8)$, is plotted. Additionally, the BER of Satcom relying on the single-mode-aided IM [14] (Sat-IM) is also provided. In the Sat-IM system, only 2 out of 4 available indices are activated for transmitting 32QAM symbols (denoted by "Sat-IM $(4, 2, 32)$"). The *matched filter bounds* of the aforementioned three schemes are also shown, serving as benchmarks. Considering that the FB equalizer of [88] exhibits the near-optimal performance in Satcom among existing literatures, the BER curve of Sat-DMIM relying on the FB equalizer (denoted by "Sat-DMIM (FB)") is shown for comparison. As seen from Fig. 3.3, in terms of *matched filter bound*, the DM-aided IM scheme outperforms its single-mode-aided and conventional single-carrier counterparts in the high-SNR region, in agreement with the findings of [44–46]. This confirms the BER performance advantage of the bits conveyed in the DM-aided IM domain [10]. In nonlinear dispersive satellite channels, the proposed BP-MF algorithm almost achieves the FB equalizer, with much lower computational complexity (see Table. II), which confirms the efficiency of the amalgamated message passing method based iterative detection algorithm. Observing that similar to the phenomenon shown in [44] for single-carrier systems over linear dispersive channels, the BER performance loss of Sat-DMIM recorded in case of free-ISI bound is lower than that of Sat-IM and Sat-8PSK over nonlinear dispersive channels. More specifically, the performance advantage of the Sat-DMIM $(4, 2, 4, 8)$ scheme over the conventional Sat-8PSK and

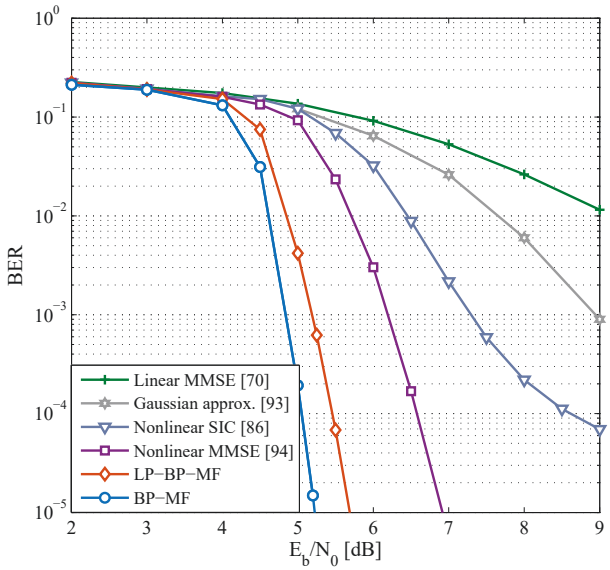

Figure 3.5: The BER performance of the proposed BP-MF and PL-BP-MF compared with existing Linear MMSE [70], Nonlinear MMSE [94], Nonlinear SIC [86] and Gaussian approx. [93] algorithms. The Sat-DMIM system (characterized by "$(4, 2, 4, 8)$") communicates over nonlinear dispersive satellite channels modeled by Volterra model "Nonlinear Channel-I" [94]. The effective throughput is 2 bits/symbol.

Sat-IM $(4, 2, 32)$ is approximately 1 dB and 1.5 dB, respectively, if the target BER is 10^{-5}. This means that using the proposed BP-MF algorithm, the Sat-DMIM scheme transmitting part of the information bits by the combination of two distinguishable constellation modes indeed succeeds in improving the performance over classical Satcom and Sat-IM with nonlinear dispersive distortions. Note that although the Sat-IM scheme also implicitly conveys additional information using the active indices in the IM domain, there exist some deactivated time slots within each subframe [14], which significantly limits the spectral efficiency of Sat-IM. Therefore, the Sat-IM employs the highest order modulation amongst the aforementioned schemes and thus exhibits the worst BER performance, since higher order modulations tend to experience stronger nonlinear distortions over satellite channels [3].

In Fig. 3.5, the BER performance of the proposed BP-MF and PL-BP-MF based iterative detection algorithms is compared to that of existing linear MMSE equalizer of [70] (denoted by "Linear MMSE"), nonlinear equalizer of [94] (denoted by "Nonlinear MMSE") and nonlinear soft interference canceler (SIC) of

[86] (denoted by "Nonlinear SIC"). Additionally, the BER curve of the Gaussian noise approximation based approach of [93] (denoted by "Gaussian approx.") is also shown for comparison. The system parameters are the same as in Fig. 3.3. It is seen that the Sat-DMIM performance degrades significantly, if the effect of nonlinear distortions is ignored. By regarding the nonlinear channels as additive Gaussian noise channel associated with a signal attenuation, the approach of [93] delivers better BER performance compared to the linear MMSE receiver. However, the performance loss imposed by this method is still considerable, because the Gaussian noise approximation substantially deviates from the system parameters especially for multi-level modulations. The nonlinear SIC of [86] improves the Sat-DMIM performance by explicitly taking part of the nonlinear ISI terms into account. We point out that the SIC method is a special case of the nonlinear MMSE equalizer of [94], where the third-order interference terms are ignored when computing the variance of the residual equalizer output error. This explains the poor BER performance of nonlinear SIC compared to nonlinear MMSE at high SNRs, where the performance is mainly limited by the nonlinear distortions instead of the additive white noise. Explicitly, it suffers from an error floor at BER= 10^{-4}. By using the nonlinear MMSE method, the BER performance improves considerably at the expense of an increase of complexity (see Table II). However, it is inferior to the proposed BP-MF and PL-BP-MF methods. One of the reason is that the nonlinear MMSE only utilizes the first- and second-order moments of data symbols, ignoring the structured *a priori* information of Sat-DMIM signal which is inherent in the DM-aided IM scheme. By contrast, the proposed BP-MF and PL-BP-MF methods take full account of the mode-selection pattern constraint and explicitly construct the structured *a priori* probabilities of Sat-DMIM signal using the *extrinsic* information provided by the SISO decoder from the previous iteration, see (3.7) and (3.8). The other reason is that the expectation of the product of two symbol vector with different indices is approximated to be zero in the nonlinear MMSE equalizer. In other words, the expectations of the parameters $\psi_{n,n}^{o\lhd\lhd}$ and $\psi_{n,n}^{\lhd\lhd o}$ are always set to be zeros and not updated with the iterations. However, both in the proposed BP-MF and PL-BP-MF algorithms, these two parameters are iteratively updated for

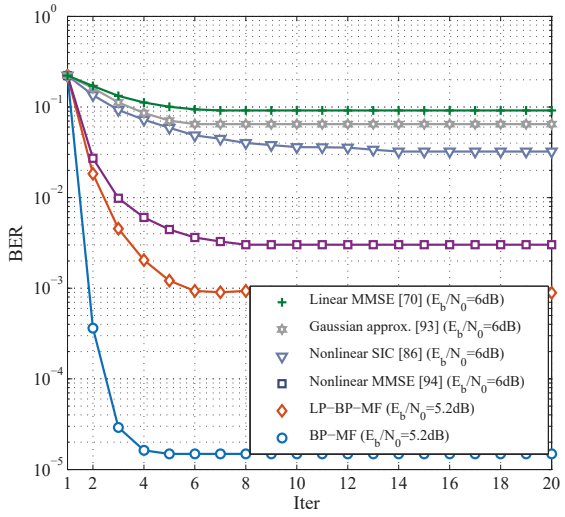

Figure 3.6: The BER performance of the proposed BP-MF and PL-BP-MF algorithms and existing methods versus iteration index. The Sat-DMIM system (characterized by "$(4, 2, 4, 8)$") communicates over nonlinear dispersive satellite channels modeled by Volterra model "Nonlinear Channel-I" [94]. The effective throughput is 2 bits/symbol.

calculating the messages passing from the variable node x_n to the observation node f_{r_m}. We also observe that compared with BP-MF, the BER performance loss of the proposed low-complexity PL-BP-MF is modest, which confirms the efficiency of the first-order Taylor series expansion approximation of the nonlinear message $\mu_{f_{r,n}\to x_n}(x_n)$ following from (3.22) and the simplification of message parameters according to (3.20) and (3.21).

Furthermore, Fig. 3.6 plots the BERs of the iterative detection algorithms as a function of the iteration index under the conditions of Fig. 3.5. The SNR is fixed to 5.2 dB and 6 dB for the proposed methods and the existing algorithms, respectively. Observe that there is an evident performance improvement, upon increasing the number of iterations. As shown in Fig. 3.6, the proposed BP-MF algorithm converges after about 5 iterations. This outcome indicates that 5 iterations on average are sufficient to meet the pre-specified stopping criterion. The convergency rate of the proposed low-complexity PL-BP-MF is slighter lower. Additionally, the convergence of the nonlinear MMSE equalizer is slow compared to the proposed BP-MF and low-complexity PL-BP-MF methods, which explains its poor BER performance at high SNRs in Fig. 3.5.

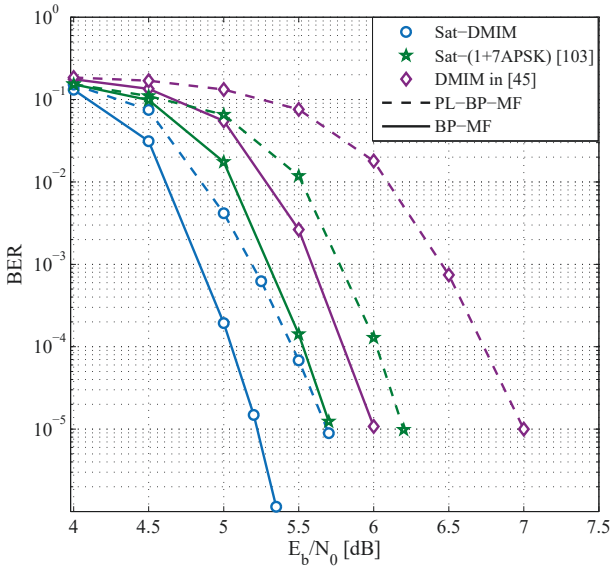

Figure 3.7: BER performance of Sat-DMIM (characterized by "$(4, 2, 4, 8)$"), $1 + 7$APSK modulated Satcom [103] and Sat-DMIM using the mode-selection mapping method of [45] communicating over nonlinear dispersive satellite channels modeled by Volterra model "Nonlinear Channel-I" [94]. Both proposed BP-MF and PL-BP-MF methods are studied for comparison.

In Fig. 3.7, the reduced-complexity PL-BP-MF iterative detection algorithm is applied to the Sat-DMIM $(4, 2, 4, 8)$ scheme and compare the BER performance attained to that of the conventional Satcom relying on an optimized $1 + 7$APSK constellation (denoted by "Sat-$(1 + 7$APSK$)$") that proposed in [103]. Additionally, the BER curve of existing DM-aided IM scheme applying the mode-selection mapping method of [45] (denoted by "DMIM in [45]") is also shown in Fig. 3.7 for comparison. It can be observed from Fig. 3.7 that the there is an evident performance improvement when applying the DMIM technique with an optimized mode-selection mapping method in nonlinear channels. Specifically, the Sat-DMIM scheme achieves about 0.5 dB better performance than the Satcom system using $1 + 7$APSK at a BER of 10^{-5}. The single-carrier system using the DMIM scheme of [45] has the worst BER performance amongst the aforementioned schemes, due to the error propagation introduced by the mapping rule applied between IM bits and mode-selection patterns. More specifically, an incorrect detection of IM bits usually results in an incorrect detection of the PSK/QAM symbols. As a result, the BER performance of the DMIM scheme of [45] is degraded especially for different

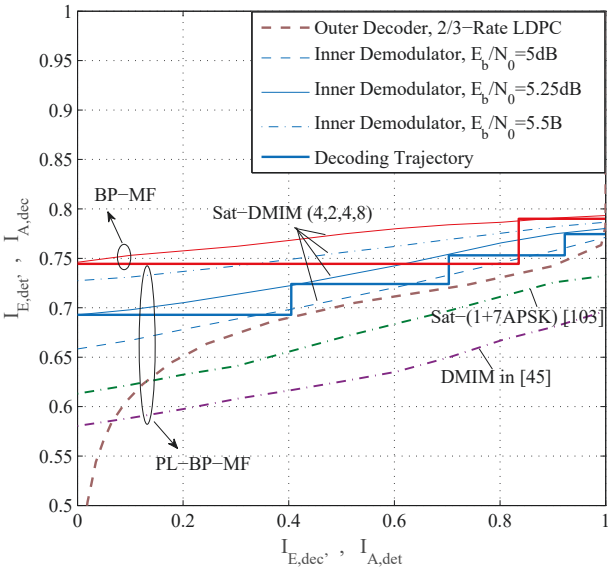

Figure 3.8: The EXIT charts of Sat-DMIM (characterized by "$(4, 2, 4, 8)$"), $1 + 7$APSK modulated Satcom [103] and Sat-DMIM relying on the proposed BP-MF and PL-BP-MF methods. The decoding trajectories are portrayed between the inner demodulator's EXIT curves and the outer LDPC decoder's EXIT curve at $E_b/N_0 = 5.25$ dB.

constellation orders in the presence of nonlinear distortions.

Fig. 3.8 shows the extrinsic information transfer chart (EXIT)[f] curves [104] of the Sat-DMIM $(4, 2, 4, 8)$ scheme's inner demodulator using the proposed PL-BP-MF method for SNR values spanning from 5 dB to = 5.5 dB with a step size of 0.25 dB. The EXIT curves of existing Sat-$(1 + 7$APSK$)$ [103] and DMIM scheme relying on the mode-selection mapping method of [45] in Fig. 3.7 are plotted at SNR= 5.5 dB for comparison. The EXIT chart of the 2/3-rate LDPC decoder is also portrayed. Additionally, Fig. 3.8 shows the Monte-Carlo simulation-based decoding trajectories of the Sat-DMIM $(4, 2, 4, 8)$ system between the inner demodulator and the outer LDPC decoder recorded at SNR= 5.25 dB for proposed BP-MF and PL-BP-MF, respectively. Observe that using the proposed PL-BP-MF method, there is no open tunnel attained between the inner curve and the outer curve at SNR= 5.5 dB for

[f]The EXIT charts allow us to analyze the convergence behavior of iteratively decoded systems, where the exchange of *extrinsic* information between the inner demodulator and the outer LDPC decoder can be visualized by calculating the mutual information between information bits and the soft values [104].

Sat-$(1+7\text{APSK})$ and DMIM scheme of [45]. By contrast, an open tunnel is seen in the Sat-DMIM $(4, 2, 4, 8)$ system for SNR beyond 5 dB. According to the open tunnel recorded at SNR= 5.25 dB, the BP-MF iterative detection algorithm is expected to converge faster than PL-BP-MF and at least 3 iterations are required for reaching the mutual information point $(1.0, 1.0)$.

3.6 Conclusions

In this chapter, the IM technique is exploited to increase transmission rates of code-aided Satcom by conveying information through two distinguishable constellation modes and their permutations. In order to combat both linear and nonlinear ISI imposed by satellite channels, an FG-based iterative detection algorithm is conceived for Satcom relying on DM-aided IM. On the FG the nonlinear correlation amongst Sat-DMIM symbols resulting from both satellite channels and the mode-selection mapping is explicitly represented. Then the amalgamated BP and MF message passing algorithm is derived over this FG for detecting both IM bits and ordinary constellation mapping bits while eliminating the linear and nonlinear distortions. The complexity of the iterative detection algorithm is further reduced by linearizing some high-order terms appearing in nonlinear messages using the *a posteriori* estimates of Sat-DMIM symbols obtained from the previous iteration. Simulation results demonstrate the effectiveness of the proposed amalgamated BP-MF-based and partial linearization approximation-based iterative detection algorithms.

Chapter 4

Iterative Detection for Faster-than-Nyquist signalling in Dispersive Channels

In Chapters 2 and 3, the iterative detection algorithms for IM-aided high spectral efficiency wireless communication systems are investigated. In this chapter, the detection of FTN systems will be studied.

4.1 Introduction

FTN signalling conveys more data symbols per unit time while preserving the same signalling bandwidth. However, the ISI imposed by FTN signalling is noticeable. Moreover, the ISI becomes more severe as the transmission rate and the frequency band increase. Therefore, advanced detection algorithms should be developed at the receiver for handling with the ISI.

The existing methods assume that FTNS are transmitted over AWGN channels or fading channels with perfectly known CSI at the receiver. Detection of FTNS in unknown frequency-selective fading channels is more challenging. In [62], a time-domain joint channel estimation and detection algorithm is proposed for FTNS in frequency-selective fading channels. However, the complexity of this approach grows cubically with the number of ISI taps. Moreover, with channel delay spreads

spanning tens or hundreds of symbol intervals in broadband wireless communications and underwater acoustic communications, etc., the CP-based [17,56,60] and overlap-based [58] algorithms suffer from severe power and spectral efficiency loss due to the large number of ISI taps resulting from both the FTNS and multipath channels.

Among the literature, there are many research results on studying the effect of multipath propagation (see e.g. [105–107]) and developing suboptimal joint channel estimation and detection methods alternative to the optimal MAP-based algorithms for Nyquist systems with low complexity (see e.g. [108–111]). Many of these methods have been extended to coded systems, where iterative process is performed between the equalizer and the SISO decoder through exchanging extrinsic information (see e.g. [112–114]). In conjunction with a SISO decoder, a class of iterative message-passing-based joint channel estimation, detection and decoding methods via FG [115] has gained a lot of attractions. In [116], the colored noise at the receiver side due to the sampling of FTN signalling is modeled by a pth-order autoregressive (AR) process, and then a time-domain linear MMSE turbo equalizer is developed based on FG. Detection of FTN signalling with imperfect synchronization is studied in [117], where a sum-product and variational message passing (SP-VMP) algorithm is developed for low-complexity FTN signalling receivers. In [63], by reformulating the problem of FTN signalling detection in doubly-selective channels into a linear state-space model, Gaussian message passing (GMP) on FG is developed and the robustness to the imperfect channel knowledge is considered. Recently, GAMP [118] is proposed to efficiently transform a linearly mixed vector estimation problem into a scalar one with low complexity. Applications of the GAMP method have been found in compressed sensing and wireless communications [75, 119, 120].

Motivated by these considerations, in this chapter, low-complexity frequency-domain joint channel estimation and decoding (JCED) algorithms are developed for FTN signalling over frequency-selective fading channels based on variational Bayesian (VB) framework [121]. By partitioning the symbol sequence into several subblocks, an alternative frequency-domain system model is derived, where the interferences from the adjacent subblocks due to the absence of CP are explicitly represented. The autocorrelation matrix of the colored noise in frequency domain

imposed by FTN signalling is approximated by a circulant matrix, which leads to trivial matrix inversions. Considering that the MAP symbol detection is intractable, the MF approximation is used to obtain the approximated posterior distributions via maximizing the inverse variational free energy (VFE), which results in iteratively updating channel estimation and data detection. Based on Gaussian approximation of data symbols, only the mean vectors and the covariance matrices have to be updated iteratively in the proposed JCED algorithm. By further exploiting the correlation between the elements in the frequency-domain CIRs, a refined JCED (R-JCED) algorithm is proposed to improve the accuracy of channel estimation. Moreover, in contrast to the Gaussian approximation of data symbols in JCED and R-JCED algorithms, the GAMP algorithm is further employed and incorporated it into the iteration between the channel estimation and decoder, which leads to two algorithms, i.e., GAMP-JCED and R-GAMP-JCED. Simulation results show that the proposed algorithms outperform the existing CP-based and overlap-based MMSE-FDE methods. Without substantially increasing transmitter energy per bit, FTN signalling based on the proposed algorithms can transmit with rate up to 67% higher than the Nyquist one, while the overall complexity is affordable in practical systems.

The main contributions of this chapter are summarized in the following:

- Instead of AWGN channels or known frequency-selective fading channels, the joint channel estimation and decoding algorithm is studied for FTN signalling in frequency-selective fading channels. A frequency-domain system model is developed without utilizing CP, where the statistics of the colored noise is approximated properly to reduce the computational complexity.

- Based on the VB framework, a joint channel estimation and decoding (JCED) algorithm and its refined version R-JCED are proposed. To further reduce the BER performance loss due to the Gaussian approximation of data symbols in JCED and R-JCED, a GAMP-based equalizer is developed and embedded to the iteration between channel estimation and decoding in a turbo receiver, named GAMP-JCED and R-GAMP-JCED.

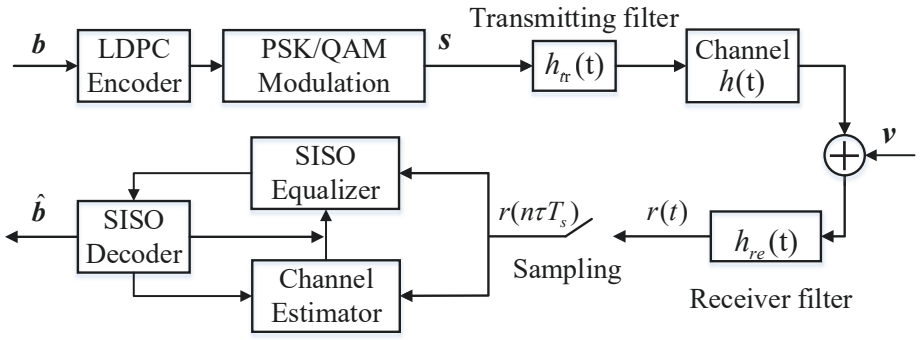

Figure 4.1: System model of the coded FTN signalling.

The rest of this chapter is organized as follows. Section 4.2 presents the system model of FTN signalling over frequency-selective fading channels. In Section 4.3, the proposed JCED, R-JCED, GAMP-JCED and R-GAMP-JCED algorithms are derived in details and their complexities are analyzed. The performance results of the proposed algorithms are evaluated via Monte Carlo simulations in Section 4.4. Conclusions are drawn in Section 4.5.

4.2 System Model

Consider a turbo FTN signalling system as depicted in Fig. 4.1. The information bits are LDPC encoded and mapped to a 2^J-PSK/QAM symbol sequence $\mathbf{s} = [s_0, \ldots, s_{K-1}]^T$, with $s_k \in \mathcal{X}$ corresponding to a length-J coded sequence $\mathbf{b}_k = [b_k^0, \ldots, b_k^{J-1}]^T$, $\mathcal{X} \triangleq \{a_0, \ldots, a_{2^J-1}\}$. The data symbols are passed through a pulse shaping filter $h_{tr}(t)$, $\int_{-\infty}^{+\infty} |h_{tr}(t)|^2 \mathrm{d}t = 1$, with symbol interval τT_s, where $0 < \tau \leqslant 1$ is the packing ratio, T_s is the symbol period of Nyquist signalling. Obviously, $\tau = 1$ corresponds to the conventional Nyquist signalling. The signal is transmitted over the time-varying frequency-selective fading channel $h(t, \iota)$. Due to the finite time support and causality, we can assume that $h(t, \iota) = 0$ for $\iota < 0$ and $\iota > \iota_{\max}$, with $\iota_{\max}$ the maximal delay spread. At the receiver side, the signal after receiver filter

$h_{re}(t) = h_{tr}^*(-t)$ can be expressed as

$$r(t) = \sum\nolimits_k s_k \int_{-\infty}^{\infty} h(t,\iota) g(t - k\tau T_s - \iota) \mathrm{d}\iota + v(t), \tag{4.1}$$

where $\star$ denotes linear convolution, $g(t) = (h_{re} \star h_{tr})(t)$, $v(t) = \int_{-\infty}^{\infty} n_0(\iota) h_{tr}^*(\iota - t) \mathrm{d}\iota$ is a Gaussian noise process with zero mean and N_0 power spectrum density. After sampling with period τT_s, we obtain the samples for the data detection and channel estimation problem, given by

$$r[n] = \sum\nolimits_k \sum\nolimits_l s_k h_{n,l} g[n - k - l] + v[n], \tag{4.2}$$

where $r[n] = r(n\tau T_s)$, $h_{n,l}$ is the channel impulse response (CIR) at time index n, $g[n] = g(n\tau T_s)$, and $v[n] = v(n\tau T_s)$. We assume that $h_{n,l}$ equals zero except for $l \in [0, L_h - 1]$, where L_h is the channel length, and $g(t) = 0$ for t outside an interval centered in 0 of size $2\nu\tau T_s$ (the value of ν is selected according to τ, see [17]). Note that $\{v[n]\}$ in (4.2) is a zero-mean colored noise process with autocorrelation function $[\mathbf{R}_v]_{nm} = E\{v[n]v^*[m]\} = N_0 g[n - m]$.

4.3 Joint Channel Estimation and Decoding Algorithms in Dispersive Channels

The symbol sequence $\mathbf{s}$ is partitioned into P length-M subblocks with $K = PM$ (P is guaranteed to be an integer by system design, $M \geq L-1$, where $L = L_h + L_g - 1$, $L_g = 2\nu + 1$). Denote the pth subblock by $\mathbf{s}_p \triangleq [s_{pM}, \ldots, s_{(p+1)M-1}]^T$, $p = 0, \ldots, P-1$. Assuming that the CIR is fixed over the duration of one subblock, the pth observation vector corresponding to $\mathbf{s}_p$ is collected as $\mathbf{r}_p \triangleq [r_{pM}, \ldots, r_{pM+N-1}]^T$, $N = M + L - 1$, which is represented by

$$\mathbf{r}_p = \mathbf{h}_p \star \mathbf{g} \star \mathbf{s}_p + \mathbf{v}_{p-1}^{tail} + \mathbf{v}_{p+1}^{head} + \mathbf{v}_p', \tag{4.3}$$

where $\mathbf{h}_p \triangleq [h_{p,0}, \ldots, h_{p,L_h-1}]^T$ with $h_{p,l}$ being the lth tap of the CIR in the pth subblock, $\mathbf{g} \triangleq [g[-\nu], \ldots, g[\nu]]^T$, $\mathbf{g} \in \mathbb{R}^{L_g \times 1}$, and $\mathbf{v}_p' \triangleq [v[pM], \ldots, v[pM + N - 1]]^T$ is the colored noise vector with autocorrelation matrix $\mathbf{R}_{\mathbf{v}_p'}$. Considering that the autocorrelation matrix $\mathbf{R}_{\mathbf{v}_p'}$ is not diagonal as that in Nyquist system, proper

approximation is required for canceling the effect of the colored noise. Note that unlike the unknown vector $\mathbf{h}_p$, the ISI taps vector $\mathbf{g}$ introduced by FTN signalling is known and remains same for all subblocks. Performance loss and complexity increase may occur if known ISI in FTN signalling is intentionally absorbed into the unknown ISI taps of the CIRs. Due to the absence of CP, the interference from the tail of the $(p-1)th$ subblock and the head of the $(p+1)th$ subblock is also considered in (4.3), where $\mathbf{v}_{p-1}^{tail} \triangleq \mathbf{A}_t(\mathbf{h}_{p-1} \star \mathbf{g} \star \mathbf{s}_{p-1})$ and $\mathbf{v}_{p+1}^{head} \triangleq \mathbf{A}_h(\mathbf{h}_{p+1} \star \mathbf{g} \star \mathbf{s}_{p+1})$, with

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{0}_M & \mathbf{0}_{M\times(L-1)} \\ \mathbf{0}_{(L-1)\times M} & \mathbf{I}_{L-1} \end{bmatrix},$$

and

$$\mathbf{A}_h = \begin{bmatrix} \mathbf{I}_{L-1} & \mathbf{0}_{(L-1)\times M} \\ \mathbf{0}_{M\times(L-1)} & \mathbf{0}_M \end{bmatrix},$$

for $p = 0, \ldots, P-1$ ($\mathbf{v}_{-1}^{tail} = 0$ and $\mathbf{v}_{P}^{head} = 0$). By appending zeros to $\mathbf{h}_p$, $\mathbf{g}$ and $\mathbf{s}_p$, Eq. (4.3) can be transformed into the following frequency-domain model

$$\mathbf{y}_p = \mathbf{c}_p \odot \underline{\mathbf{g}} \odot \mathbf{t}_p + \mathbf{w}_{p-1}^{tail} + \mathbf{w}_{p+1}^{head} + \mathbf{w}'_p, \tag{4.4}$$

where $\odot$ denotes element-wise product, $\mathbf{y}_p = \mathbf{F}_N\mathbf{r}_p$, $\mathbf{w}_{p-1}^{tail} = \mathbf{F}_N\mathbf{v}_{p-1}^{tail}$, $\mathbf{w}_{p+1}^{head} = \mathbf{F}_N\mathbf{v}_{p+1}^{head}$, $\mathbf{w}'_p = \mathbf{F}_N\mathbf{v}'_p$, $\underline{\mathbf{g}} = \sqrt{N}\mathbf{F}_N\mathbf{T}_2\mathbf{g}$, and

$$\mathbf{c}_p = \sqrt{N}\mathbf{F}_N\mathbf{T}_1\mathbf{h}_p, \tag{4.5}$$

$$\mathbf{t}_p = \mathbf{F}_N\mathbf{T}_3\mathbf{s}_p = \mathbf{F}_N\underline{\mathbf{s}}_p, \tag{4.6}$$

with $\mathbf{T}_1 = [\mathbf{I}_{L_h}\ \mathbf{0}_{L_h\times(N-L_h)}]^T$, $\mathbf{T}_2 = [\mathbf{I}_{L_g}\ \mathbf{0}_{L_g\times(N-L_g)}]^T$, and $\mathbf{T}_3 = [\mathbf{I}_M\ \mathbf{0}_{M\times(N-M)}]^T$. The autocorrelation matrix $\mathbf{R}_{\mathbf{w}'_p}$ of $\mathbf{w}'_p$ can be derived as

$$\begin{aligned} \mathbf{R}_{\mathbf{w}'_p} &= \mathbf{F}_N\mathbf{R}_{\mathbf{v}'_p}\mathbf{F}_N^H = N_0\left(\mathbf{F}_N\mathbf{R}''\mathbf{F}_N^H - \mathbf{F}_N\mathbf{R}_{\triangle}\mathbf{F}_N^H\right) \\ &= N_0\left[\mathcal{D}(\sqrt{N}\mathbf{F}_N\boldsymbol{\lambda}'') - \mathbf{F}_N\mathbf{R}_{\triangle}\mathbf{F}_N^H\right] \\ &\approx N_0\left[\mathcal{D}(\sqrt{N}\mathbf{F}_N\boldsymbol{\lambda}'') - \mathcal{D}(\mathbf{F}_N\mathbf{R}_{\triangle}\mathbf{F}_N^H)\right] \triangleq \boldsymbol{\Lambda}'_p, \end{aligned} \tag{4.7}$$

where $\mathbf{R}''=\mathbf{F}_N^H\mathcal{D}(\sqrt{N}\mathbf{F}_N\boldsymbol{\lambda}'')\mathbf{F}_N$ is an $N\times N$ circulant matrix constructing from the DFT of $\boldsymbol{\lambda}'' = [g[0], \ldots, g[\nu], 0, \ldots, 0,$

$g[-\nu], \ldots, g[-1]]^T \in \mathbb{R}^{N\times 1}$, and

$$\mathbf{R}_{\triangle} = \begin{bmatrix} & & g[\nu] \ldots g[1] \\ & & \ddots \ \vdots \\ & & g[\nu] \\ & \mathbf{0} & \\ g[-\nu] & & \\ \vdots \ \ddots & & \\ g[-1] \ldots g[-\nu] & & \end{bmatrix} \in \mathbb{R}^{N\times N}.$$

The approximation error in (4.7) is acceptable when the values in $[g[-\nu], \ldots, g[-1], g[1], \ldots, g[\nu]]^T$ are small enough with large packing ratios for the FTN signalling. Note that there is no approximation error for Nyquist signalling since $\mathbf{R}_{\triangle}$ becomes a zero matrix in that case, and the approximation error also gets smaller as N increases [122]. In this way, the matrix $\mathbf{R}_{\mathbf{w}'_p}$ is approximated by a diagonal matrix denoted as $\mathbf{\Lambda}'_p$, whose diagonal elements $[\Lambda'_{pM}, \ldots, \Lambda'_{pM+N-1}]^T \triangleq \boldsymbol{\lambda}'_p$.

For brevity, we define $\mathbf{w}_p \triangleq \mathbf{w}^{tail}_{p-1} + \mathbf{w}^{head}_{p+1} + \mathbf{w}'_p$, with the mean vector and covariance matrix given as

$$\begin{aligned} E(\mathbf{w}_p) &= E(\mathbf{w}^{tail}_{p-1}) + E(\mathbf{w}^{head}_{p+1}), \\ V(\mathbf{w}_p) &\approx \lambda \mathbf{I}_N + \mathbf{\Lambda}'_p = \mathbf{\Lambda}_p, \end{aligned} \tag{4.8}$$

where $\mathbf{\Lambda}_p = \mathcal{D}([\Lambda_{pM}, \ldots, \Lambda_{pM+N-1}]) = \mathcal{D}(\boldsymbol{\lambda}'_p + \lambda \mathbf{1}_{N\times 1})$, and[a] the iterations.

$$\lambda = \frac{1}{N} \sum\nolimits_{l=0}^{L_h-1} \left[v'_{p-1} l |E(h^l_{p-1})|^2 + v'_{p+1}(L_h - 1 - l)|E(h^l_{p+1})|^2 \right], \tag{4.9}$$

where v'_p corresponds to the variance of the data estimate as shown in Section III-B[b]. In (4.8), the mean vectors of $\mathbf{w}^{tail}_{p-1}$ and $\mathbf{w}^{head}_{p+1}$ are calculated as

$$\begin{aligned} E(\mathbf{w}^{tail}_{p-1}) &= \mathbf{F}_N \Big\{ \mathbf{A}_t \big\{ \mathbf{F}^H_N [E(\mathbf{c}_{p-1}) \odot E(\underline{\mathbf{t}}_{p-1})] \big\} \Big\}, \\ E(\mathbf{w}^{head}_{p+1}) &= \mathbf{F}_N \Big\{ \mathbf{A}_h \big\{ \mathbf{F}^H_N [E(\mathbf{c}_{p+1}) \odot E(\underline{\mathbf{t}}_{p+1})] \big\} \Big\}, \end{aligned} \tag{4.10}$$

where $\underline{\mathbf{t}}_p \triangleq \mathbf{G}\mathbf{t}_p$ with $\mathbf{G} \triangleq \mathcal{D}(\underline{\mathbf{g}})$. Note that Eq. (4.10) can be efficiently performed by fast Fourier transform (FFT).

[a] In (4.9), the value of v'_p corresponds to the variance of the data estimates, which will approach to 0 with the iterations as shown in Section III-B. In other words, the approximation error of (4.8) gradually decreases along with

[b] The equations (4.8) and (4.9) are derived by applying the iterated expectation theorem [123, ch. 14]

4.3.1 JCED Based on the VB Framework

Based on the frequency-domain model in (4.4), an optimal MAP symbol detector for s_n, $n = pM, \ldots, pM + M - 1$, in the pth subblock is formulated as

$$\begin{aligned}\hat{s}_n^{\mathrm{MAP}} &= \arg\max_{s_n} p(s_n|\mathbf{y}_p) \\ &= \arg\max_{s_n} \sum_{\mathbf{s}_p \backslash s_n} \int_{\mathbf{h}_p} p(\mathbf{y}_p|\mathbf{s}_p, \mathbf{h}_p) p(\mathbf{h}_p) p(\mathbf{s}_p) \mathrm{d}\mathbf{h}_p \\ &= \arg\max_{s_n} \sum_{\mathbf{s}_p \backslash s_n} p(\mathbf{y}_p|\mathbf{s}_p) p(\mathbf{s}_p). \end{aligned} \tag{4.11}$$

Using the iterated expectation theorem [123, ch. 14] and its analog in covariance, we obtain the mean vector and the covariance matrix of the pdf $p(\mathbf{y}_p|\mathbf{s}_p) \propto \mathcal{N}_c(\mathbf{y}_p; E(\mathbf{y}_p|\mathbf{s}_p), V(\mathbf{y}_p|\mathbf{s}_p))$ by

$$\begin{aligned} E(\mathbf{y}_p|\mathbf{s}_p) &= \mathbf{G}\mathcal{D}(\mathbf{F}_N \mathbf{T}_3 \mathbf{s}_p)\sqrt{N}\mathbf{F}_N \mathbf{T}_1 E(\mathbf{h}_p) + E(\mathbf{w}_p), \\ V(\mathbf{y}_p|\mathbf{s}_p) &= \mathbf{G}\mathcal{D}(\mathbf{F}_N \mathbf{T}_3 \mathbf{s}_p) N \mathbf{F}_N \mathbf{T}_1 V(\mathbf{h}_p) \mathbf{T}_1^H \mathbf{F}_N^H \mathcal{D}^H(\mathbf{F}_N \mathbf{T}_3 \mathbf{s}_p) \mathbf{G}^H + \mathbf{\Lambda}_p, \end{aligned} \tag{4.12}$$

where $E(\mathbf{h}_p)$ and $V(\mathbf{h}_p)$ are the mean vector and covariance matrix of $\mathbf{h}_p$. By substituting (4.12) into (4.11), it is observed that there is no closed-form solution to the maximization problem, and the search over the symbol space is required.

To tackle this problem, a VB framework is used to approximate the joint posterior distribution $p(\mathbf{s}_p, \mathbf{h}_p|\mathbf{y}_p)$ with an auxiliary function $Q(\mathbf{s}_p, \mathbf{h}_p)$ and then find $\mathbf{s}_p$ and $\mathbf{h}_p$ to maximize $Q(\mathbf{s}_p, \mathbf{h}_p)$. Further complexity reduction can be realized via mean-field (MF) approximation, i.e., $Q(\mathbf{s}_p, \mathbf{h}_p) = q(\mathbf{s}_p)q(\mathbf{h}_p)$. This approximation implicity adopts the assumption that $\mathbf{s}_p$ and $\mathbf{h}_p$ are independent given the observed vector $\mathbf{y}_p$. Note that the marginalization of $q(\mathbf{s}_p)$ for detection is much simpler than that in (4.11), as long as the auxiliary posterior pdf $q(\mathbf{s}_p)$ is chosen properly. Since $\arg\max_{\mathbf{s}_p} \int Q(\mathbf{s}_p, \mathbf{h}_p) d\mathbf{h}_p = \arg\max_{\mathbf{s}_p} q(\mathbf{s}_p)$, the complex integral calculation problem is avoided. In this framework, we first calculate the log-likelihood function $\ln p(\mathbf{y}_p)$ by

$$\begin{aligned} \ln p(\mathbf{y}_p) &= \ln \sum\nolimits_{\mathbf{s}_p} \int_{\mathbf{h}_p} p(\mathbf{s}_p, \mathbf{h}_p, \mathbf{y}_p) \mathrm{d}\mathbf{h}_p \\ &\geq \sum\nolimits_{\mathbf{s}_p} \int_{\mathbf{h}_p} q(\mathbf{s}_p) q(\mathbf{h}_p) \ln \frac{p(\mathbf{s}_p, \mathbf{h}_p, \mathbf{y}_p)}{q(\mathbf{s}_p) q(\mathbf{h}_p)} \mathrm{d}\mathbf{h}_p \\ &\triangleq \mathcal{F}_{\mathcal{H}}, \end{aligned} \tag{4.13}$$

where the equality holds if and only if $q(\mathbf{s}_p)q(\mathbf{h}_p) = p(\mathbf{s}_p, \mathbf{h}_p|\mathbf{y}_p)$ [121]. This means that minimizing the KL-divergence $\mathrm{KL}[q(\mathbf{s}_p)q(\mathbf{h}_p)||p(\mathbf{s}_p, \mathbf{h}_p|\mathbf{y}_p)]$ is equivalent to maximizing the inverse variational free energy $\mathcal{F}_{\mathcal{H}}$. In this way, the problem of finding the auxiliary function $Q(\mathbf{s}_p, \mathbf{h}_p)$ which is a close approximation to *a posteriori* density function $p(\mathbf{s}_p, \mathbf{h}_p|\mathbf{y}_p)$ is transformed into finding $q(\mathbf{s}_p)$ and $q(\mathbf{h}_p)$ to maximize $\mathcal{F}_{\mathcal{H}}$.

By taking functional derivatives of $\mathcal{F}_{\mathcal{H}}$ w.r.t. $q(\mathbf{s}_p)$ and $q(\mathbf{h}_p)$ alternatively, the following updates are obtained in the γth iteration

$$\ln q^{\gamma}(\mathbf{h}_p) \propto \mathrm{E}_{q^{\gamma-1}(\mathbf{s}_p)}\left[\ln p(\mathbf{y}_p|\mathbf{s}_p, \mathbf{h}_p)\right] + \ln p(\mathbf{h}_p), \tag{4.14}$$

$$\ln q^{\gamma}(\mathbf{s}_p) \propto \mathrm{E}_{q^{\gamma}(\mathbf{h}_p)}\left[\ln p(\mathbf{y}_p|\mathbf{s}_p, \mathbf{h}_p)\right] + \ln p(\mathbf{s}_p), \tag{4.15}$$

where $\mathrm{E}_{q^{\gamma}(\mathbf{a})}[\mathbf{A}]$ denotes the expectation of $\mathbf{A}$ w.r.t. $\mathbf{a}$ given the condition function $q^{\gamma}(\mathbf{a})$. Note that different from directly extending the conventional methods for Nyquist signalling to FTN systems, the ISI taps to be estimated here are simply from the unknown time-dispersive channels instead of the combinations that contains the known ISI in FTN signalling.

With the Gaussian assumptions of the distributions of time-domain symbols and CIRs, the *a priori* pdfs[c] are represented by

$$p(\mathbf{h}_p) \propto \mathcal{N}_c(\mathbf{h}_p; \bar{\mathbf{m}}_{\mathbf{h}_p}, \bar{\mathbf{V}}_{\mathbf{h}_p}), \tag{4.16}$$

$$p(\mathbf{s}_p) \propto \mathcal{N}_c(\mathbf{s}_p; \bar{\mathbf{m}}_{\mathbf{s}_p}, \bar{\mathbf{V}}_{\mathbf{s}_p}) \approx \mathcal{N}_c(\mathbf{s}_p; \bar{\mathbf{m}}_{\mathbf{s}_p}, \bar{\beta}_{\mathbf{s}_p}\mathbf{I}_M), \tag{4.17}$$

where $\bar{\beta}_{\mathbf{s}_p} = \frac{1}{M}\mathrm{tr}(\bar{\mathbf{V}}_{\mathbf{s}_p})$. Note that this approximation has been extensively employed in the design of frequency-domain equalizers [110, 111, 124]. The updating process of this *a priori* information will be discussed later in this section. Given the Gaussian priors, the posterior pdfs $q(\mathbf{s}_p)$ and $q(\mathbf{h}_p)$ are Gaussian functions, which are denoted as $q(\mathbf{s}_p) \propto \mathcal{N}_c(\mathbf{s}_p; \hat{\mathbf{m}}_{\mathbf{s}_p}, \hat{\mathbf{V}}_{\mathbf{s}_p})$ and $q(\mathbf{h}_p) \propto \mathcal{N}_c(\mathbf{h}_p; \hat{\mathbf{m}}_{\mathbf{h}_p}, \hat{\mathbf{V}}_{\mathbf{h}_p})$.

[c]The *a priori* pdf of the symbol sequence is assumed to be fully factorized, since pseudorandom interleavers are employed in a turbo process. However, the interleavers and the deinterleavers are omitted in Fig. 4.1, because a sufficiently long pseudo-random LDPC code is used in our concerned system.

By substituting the Gaussian prior pdf $p(\mathbf{h}_p)$ into (4.14), we have

$$\begin{aligned} q^{\gamma}(\mathbf{h}_p) \propto \exp\Big\{ &-\Big[\hat{\mathbf{z}}_p - \mathcal{D}(\hat{\mathbf{m}}_{\underline{\mathbf{t}}_p})\sqrt{N}\mathbf{F}_N\mathbf{T}_1\mathbf{h}_p\Big]^H \hat{\mathbf{\Lambda}}_p^{-1}\Big[\hat{\mathbf{z}}_p \\ &-\mathcal{D}(\hat{\mathbf{m}}_{\underline{\mathbf{t}}_p})\sqrt{N}\mathbf{F}_N\mathbf{T}_1\mathbf{h}_p\Big] - N\mathbf{h}_p^H\mathbf{T}_1^H\mathbf{F}_N^H\mathcal{D}(\hat{\mathbf{V}}_{\underline{\mathbf{t}}_p})\hat{\mathbf{\Lambda}}_p^{-1} \\ &\times \mathbf{F}_N\mathbf{T}_1\mathbf{h}_p - \left[\mathbf{h}_p - \bar{\mathbf{m}}_{\mathbf{h}_p}\right]^H \bar{\mathbf{V}}_{\mathbf{h}_p}^{-1}\left[\mathbf{h}_p - \bar{\mathbf{m}}_{\mathbf{h}_p}\right]\Big\}, \end{aligned} \tag{4.18}$$

where $\hat{\mathbf{z}}_p \triangleq \mathbf{y}_p - E(\mathbf{w}_p)$, $\hat{\mathbf{\Lambda}}_p \triangleq \mathbf{\Lambda}'_p + \hat{\lambda}\mathbf{I}_N$, $\hat{\mathbf{m}}_{\underline{\mathbf{t}}_p} \triangleq \mathbf{G}\hat{\mathbf{m}}_{\mathbf{t}_p}$, and $\hat{\mathbf{V}}_{\underline{\mathbf{t}}_p} \triangleq \mathbf{G}\hat{\mathbf{V}}_{\mathbf{t}_p}\mathbf{G}^H$. Note the matrix $\mathbf{G}$ is corresponding to the frequency-domain ISI information in the FTN signalling, which is known and depends only on the transmitting filter. In (4.18), $\{\hat{\mathbf{z}}_p, \hat{\mathbf{\Lambda}}_p\}$, and $\{\hat{\mathbf{m}}_{\underline{\mathbf{t}}_p}, \hat{\mathbf{V}}_{\underline{\mathbf{t}}_p}\}$ are obtained from the $(\gamma-1)th$ iteration and the subscript "$\gamma-1$" is dropped for simplicity of notation. After some direct manipulations, we obtain the posterior pdf $q^{\gamma}(\mathbf{h}_p)$ which is proportional to $\mathcal{N}_c(\mathbf{h}_p; \hat{\mathbf{m}}_{\mathbf{h}_p}, \hat{\mathbf{V}}_{\mathbf{h}_p})$ at the γth iteration with

$$\begin{aligned} \hat{\mathbf{V}}_{\mathbf{h}_p} &= \left[\check{\mathbf{V}}_{\mathbf{h}_p}^{-1} + \bar{\mathbf{V}}_{\mathbf{h}_p}^{-1}\right]^{-1}, \\ \hat{\mathbf{m}}_{\mathbf{h}_p} &= \hat{\mathbf{V}}_{\mathbf{h}_p}\left[\check{\mathbf{V}}_{\mathbf{h}_p}^{-1}\check{\mathbf{m}}_{\mathbf{h}_p} + \bar{\mathbf{V}}_{\mathbf{h}_p}^{-1}\bar{\mathbf{m}}_{\mathbf{h}_p}\right], \end{aligned} \tag{4.19}$$

where

$$\check{\mathbf{V}}_{\mathbf{h}_p} = \Big[N\mathbf{T}_1^H\mathbf{F}_N^H\big[\mathcal{D}^H(\hat{\mathbf{m}}_{\underline{\mathbf{t}}_p})\hat{\mathbf{\Lambda}}_p^{-1}\mathcal{D}(\hat{\mathbf{m}}_{\underline{\mathbf{t}}_p}) + \mathcal{D}(\hat{\mathbf{V}}_{\underline{\mathbf{t}}_p})\hat{\mathbf{\Lambda}}_p^{-1}\big]\mathbf{F}_N\mathbf{T}_1\Big]^{-1} \tag{4.20}$$

$$\check{\mathbf{m}}_{\mathbf{h}_p} = \check{\mathbf{V}}_{\mathbf{h}_p}\left[\sqrt{N}\mathbf{T}_1^H\mathbf{F}_N^H\mathcal{D}^H(\hat{\mathbf{m}}_{\underline{\mathbf{t}}_p})\hat{\mathbf{\Lambda}}_p^{-1}\hat{\mathbf{z}}_p\right]. \tag{4.21}$$

Note that the calculation in (4.20) involves an N-dimensional matrix inversion, which requires some additional operations to reduce the computational complexity. To avoid performance loss, the direct diagonalization of $\check{\mathbf{V}}_{\mathbf{h}_p}$ cannot be employed [125], since it may damage the received information observed in (4.21). To derive the appropriate approximation, we rewrite $\hat{\mathbf{V}}_{\mathbf{h}_p}$ as

$$\begin{aligned} \hat{\mathbf{V}}_{\mathbf{h}_p} &\approx \mathbf{T}_1^H\left(N\mathbf{F}_N^H\mathbf{W}_{\mathbf{h}_p}\mathbf{F}_N + \mathbf{T}_1\bar{\mathbf{V}}_{\mathbf{h}_p}^{-1}\mathbf{T}_1^H\right)^{-1}\mathbf{T}_1 \\ &= \mathbf{T}_1^H\mathbf{F}_N^H\left(N\mathbf{W}_{\mathbf{h}_p} + \mathbf{F}_N\mathbf{T}_1\bar{\mathbf{V}}_{\mathbf{h}_p}^{-1}\mathbf{T}_1^H\mathbf{F}_N^H\right)^{-1}\mathbf{F}_N\mathbf{T}_1 \end{aligned} \tag{4.22}$$

where $\mathbf{W}_{\mathbf{h}_p} = \mathcal{D}^H(\hat{\mathbf{m}}_{\underline{\mathbf{t}}_p})\hat{\mathbf{\Lambda}}_p^{-1}\mathcal{D}(\hat{\mathbf{m}}_{\underline{\mathbf{t}}_p}) + \mathcal{D}(\hat{\mathbf{V}}_{\underline{\mathbf{t}}_p})\hat{\mathbf{\Lambda}}_p^{-1}$ is a diagonal matrix. In this way, as long as the matrix $\mathbf{F}_N\mathbf{T}_1\bar{\mathbf{V}}_{\mathbf{h}_p}^{-1}\mathbf{T}_1^H\mathbf{F}_N^H$ is approximated to be diagonal, i.e., $\mathbf{F}_N\mathbf{T}_1\bar{\mathbf{V}}_{\mathbf{h}_p}^{-1}\mathbf{T}_1^H\mathbf{F}_N^H \approx \bar{\beta}_{\mathbf{h}_p}^{-1}\mathbf{I}_N$ with $\bar{\beta}_{\mathbf{h}_p} = \frac{1}{L_h}\mathrm{tr}(\bar{\mathbf{V}}_{\mathbf{h}_p})$, all matrix inversions are trivial.

As a result, $\hat{\mathbf{V}}_{\mathbf{h}_p}$ can be approximated as

$$\hat{\mathbf{V}}_{\mathbf{h}_p} \approx \mathbf{T}_1^H \mathbf{F}_N^H \left(N\mathbf{W}_{\mathbf{h}_p} + L_h \left(\mathrm{tr}\left(\bar{\mathbf{V}}_{\mathbf{h}_\mathrm{p}} \right) \right)^{-1} \mathbf{I}_N \right)^{-1} \mathbf{F}_N \mathbf{T}_1 \tag{4.23}$$

$$\approx \underbrace{\frac{1}{N}\mathrm{tr}\left[\left(N\mathbf{W}_{\mathbf{h}_p} + L_h \left(\mathrm{tr}\left(\bar{\mathbf{V}}_{\mathbf{h}_\mathrm{p}} \right) \right)^{-1} \mathbf{I}_N \right)^{-1} \right]}_{\hat{\beta}_{\mathbf{h}_p}} \mathbf{I}_{L_h}. \tag{4.24}$$

It will be shown that this diagonal approximation of $\hat{\mathbf{V}}_{\mathbf{h}_p}$ is of the essence in the following calculations to update the channel estimation in the current subblock utilizing the information from the previous and latter subblocks. Given the approximation of the posterior covariance matrix in (4.23), the posterior mean vector $\hat{\mathbf{m}}_{\mathbf{h}_p}$ is approximated as

$$\begin{aligned}\hat{\mathbf{m}}_{\mathbf{h}_p} &\approx \mathbf{T}_1^H \mathbf{F}_N^H \left(N\mathbf{W}_{\mathbf{h}_p} + \bar{\beta}_{\mathbf{h}_p}^{-1} \mathbf{I}_N \right)^{-1} \left[\sqrt{N} \mathcal{D}^H(\hat{\mathbf{m}}_{\underline{\mathbf{t}}_p}) \hat{\mathbf{\Lambda}}_p^{-1} \hat{\mathbf{z}}_p + \bar{\beta}_{\mathbf{h}_p}^{-1} \mathbf{F}_N \mathbf{T}_1 \bar{\mathbf{m}}_{\mathbf{h}_p} \right] \\ &\approx \sqrt{N} \mathbf{T}_1^H \mathbf{F}_N^H \left(N\mathbf{W}_{\mathbf{h}_p} + \bar{\beta}_{\mathbf{h}_p}^{-1} \mathbf{I}_N \right)^{-1} \mathcal{D}^H(\hat{\mathbf{m}}_{\underline{\mathbf{t}}_p}) \hat{\mathbf{\Lambda}}_p^{-1} \hat{\mathbf{z}}_p + \frac{\hat{\beta}_{\mathbf{h}_p}}{\bar{\beta}_{\mathbf{h}_p}} \bar{\mathbf{m}}_{\mathbf{h}_p}, \end{aligned} \tag{4.25}$$

where the approximation in the first line is $\mathbf{F}_N \mathbf{T}_1 \mathbf{T}_1^H \mathbf{F}_N^H \approx \mathbf{I}_N$, and in the second line reduces the computational complexity from $\mathcal{O}(2N \log N)$ to $\mathcal{O}(N \log N)$ based on (4.23) and (4.24).

According to the turbo principle, the *a posteriori* probability $q^{\gamma-1}(\mathbf{h}_p)$ cannot be substituted to (4.14) as the prior of $\mathbf{h}_p$. To this end, we extract this *a priori* probability from the $(p-1)$th and the $(p+1)$th subblocks by utilizing the temporal channel correlation. Based on the wide sense stationary uncorrelated scattering (WSSUS) channel model, the statistics of the channel taps can be represented by $E[h_{p,l}(h_{p+i,l'})^*] = \varphi_l \delta(l-l') J_0(2\pi f_D M\tau T_s i)$, where f_D is the Doppler spread, J_0 is the zeroth-order Bessel function of the first kind, and φ_l is the power of the lth channel tap. Therefore, the time variation of the channel can be approximately modeled by the following first-order AR process:

$$\mathbf{h}_p = \mathbf{A}\mathbf{h}_{p-1} + \mathbf{v}_h, \tag{4.26}$$

where $\mathbf{A} = J_0(2\pi f_D M\tau T_s)\mathbf{I}_{L_h}$ is the correlation coefficient matrix, $\mathbf{R} = [1 - J_0^2(2\pi f_D M\tau T_s)]\mathcal{D}(\boldsymbol{\psi})$ is the covariance matrix of $\mathbf{v}_h$, with $\boldsymbol{\psi} = [\varphi_0, \ldots, \varphi_{L_h-1}]^T$ the power delay profile. Note that an alternative method to optimize the AR parameters

by minimizing the mean square error (MSE) of the channel estimates could be able to enhance the channel estimation performance [126].

Given (4.26), the bidirectional channel estimation method involving a forward-backward recursion process is summarized in Algorithm 7.

In this way, by extracting the *a priori* probability from adjacent subblocks, the posterior pdf $q^{\gamma}(\mathbf{h}_p)$ is finally updated by (4.27). Since the matrices $\mathbf{A}$, $\mathbf{R}$, $\hat{\mathbf{V}}_{\mathbf{h}_p}$ are all diagonal, the calculations involved in the forward-backward recursion process of the channel estimation are all trival. Further from (4.5), the posterior pdf $q^{\gamma}(\mathbf{c}_p)$ at the γth iteration is finally updated by

$$\begin{aligned}\hat{\mathbf{V}}_{\mathbf{c}_p} &= N\mathbf{F}_N\mathbf{T}_1\hat{\mathbf{V}}_{\mathbf{h}_p}\mathbf{T}_1^H\mathbf{F}_N^H \approx \mathrm{tr}(\hat{\mathbf{V}}_{\mathbf{h}_\mathrm{p}})\mathbf{I}_\mathrm{N},\\ \hat{\mathbf{m}}_{\mathbf{c}_p} &= \sqrt{N}\mathbf{F}_N\mathbf{T}_1\hat{\mathbf{m}}_{\mathbf{h}_p}.\end{aligned} \tag{4.28}$$

Similarly, by substituting the Gaussian prior $p(\mathbf{s}_p)$ in (4.17) obtained from channel decoder into (4.15), the covariance matrix and the mean vector of the posterior pdf $q^{\gamma}(\mathbf{s}_p)$ can be updated by

$$\begin{aligned}\hat{\mathbf{V}}_{\mathbf{s}_p} &= \left[\mathbf{T}_3^H\mathbf{F}_N^H\left(\mathcal{D}^H(\hat{\mathbf{m}}_{\underline{\mathbf{c}}_p})\hat{\mathbf{\Lambda}}_p^{-1}\mathcal{D}(\hat{\mathbf{m}}_{\underline{\mathbf{c}}_p}) + \mathcal{D}(\hat{\mathbf{V}}_{\underline{\mathbf{c}}_p})\hat{\mathbf{\Lambda}}_p^{-1}\right)\mathbf{F}_N\mathbf{T}_3 + \bar{\mathbf{V}}_{\mathbf{s}_p}^{-1}\right]^{-1},\\ \hat{\mathbf{m}}_{\mathbf{s}_p} &= \hat{\mathbf{V}}_{\mathbf{s}_p}\left[\mathbf{T}_3^H\mathbf{F}_N^H\mathcal{D}^H(\hat{\mathbf{m}}_{\underline{\mathbf{c}}_p})\hat{\mathbf{\Lambda}}_p^{-1}\hat{\mathbf{z}}_p + \bar{\mathbf{V}}_{\mathbf{s}_p}^{-1}\bar{\mathbf{m}}_{\mathbf{s}_p}\right],\end{aligned} \tag{4.29}$$

which can be further approximated by[d]

$$\begin{aligned}\hat{\mathbf{V}}_{\mathbf{s}_p} &\approx \frac{1}{N}\mathrm{tr}\left(\bar{\beta}_{\mathbf{s}_\mathrm{p}}^{-1}\mathbf{I}_\mathrm{N} + \mathbf{W}_{\mathbf{s}_\mathrm{p}}\right)^{-1}\mathbf{I}_\mathrm{M} = \hat{\beta}_{\mathbf{s}_\mathrm{p}}\mathbf{I}_\mathrm{M},\\ \hat{\mathbf{m}}_{\mathbf{s}_p} &\approx \frac{\hat{\beta}_{\mathbf{s}_p}}{\bar{\beta}_{\mathbf{s}_p}}\bar{\mathbf{m}}_{\mathbf{s}_p} + \mathbf{T}_3\mathbf{F}_N^H\left(\bar{\beta}_{\mathbf{s}_p}^{-1}\mathbf{I}_N + \mathbf{W}_{\mathbf{s}_p}\right)^{-1}\mathcal{D}^H(\hat{\mathbf{m}}_{\underline{\mathbf{c}}_p})\hat{\mathbf{\Lambda}}_p^{-1}\hat{\mathbf{z}}_p,\end{aligned} \tag{4.30}$$

with $\mathbf{W}_{\mathbf{s}_p} \triangleq \mathcal{D}^H(\hat{\mathbf{m}}_{\underline{\mathbf{c}}_p})\hat{\mathbf{\Lambda}}_p^{-1}\mathcal{D}(\hat{\mathbf{m}}_{\underline{\mathbf{c}}_p}) + \mathcal{D}(\hat{\mathbf{V}}_{\underline{\mathbf{c}}_p})\hat{\mathbf{\Lambda}}_p^{-1}$, $\hat{\mathbf{m}}_{\underline{\mathbf{c}}_p} \triangleq \mathbf{G}\hat{\mathbf{m}}_{\mathbf{c}_p}$ and $\hat{\mathbf{V}}_{\underline{\mathbf{c}}_p} \triangleq \mathbf{G}\hat{\mathbf{V}}_{\mathbf{c}_p}\mathbf{G}^H$. The posterior pdf $q^{\gamma}(\mathbf{t}_p)$ is approximated as $\mathcal{N}_c(\mathbf{t}_p; \hat{\mathbf{m}}_{\mathbf{t}_p}, \hat{\mathbf{V}}_{\mathbf{s}_p})$, with

$$\begin{aligned}\hat{\mathbf{V}}_{\mathbf{t}_p} &= \mathbf{F}_N\mathbf{T}_3\hat{\mathbf{V}}_{\mathbf{s}_p}\mathbf{T}_3^H\mathbf{F}_N^H \approx \hat{\beta}_{\mathbf{s}_p}\mathbf{I}_N,\\ \hat{\mathbf{m}}_{\mathbf{t}_p} &= \mathbf{F}_N\mathbf{T}_3\hat{\mathbf{m}}_{\mathbf{s}_p}.\end{aligned} \tag{4.31}$$

Finally, from (4.28) and (4.31), $\hat{\mathbf{z}}_p$ and $\hat{\mathbf{\Lambda}}_p$ at the γth iteration are updated by

$$\begin{aligned}\hat{\mathbf{z}}_p &= \mathbf{y}_p - E(\mathbf{w}_p),\\ \hat{\mathbf{\Lambda}}_p &= V(\mathbf{w}_p),\end{aligned} \tag{4.32}$$

[d]The equation (4.30) can be obtained by using the approximation $\mathbf{F}_N\mathbf{T}_3\mathbf{T}_3^H\mathbf{F}_N^H \approx \mathbf{I}_N$ to (4.29).

Algorithm 7 Forward-backward recursion for channel estimation.

• **Forward recursion:**

1) Initialization: Set $\bar{\mathbf{V}}_{\mathbf{h}_0} = \infty$ and $\bar{\mathbf{m}}_{\mathbf{h}_0} = \mathbf{0}_{L_h \times 1}$.

2) For $p = 0, \ldots, P-1$

a) Substitute $\{\bar{\mathbf{V}}_{\mathbf{h}_p}, \bar{\mathbf{m}}_{\mathbf{h}_p}\}$ into (22) and (4.25), to obtain $\{\hat{\mathbf{V}}_{\mathbf{h}_p}, \hat{\mathbf{m}}_{\mathbf{h}_p}\}$.

b) Set $\hat{\mathbf{V}}^f_{\mathbf{h}_p} = \hat{\mathbf{V}}_{\mathbf{h}_p}$, and $\hat{\mathbf{m}}^f_{\mathbf{h}_p} = \hat{\mathbf{m}}_{\mathbf{h}_p}$.

c) $\bar{\mathbf{V}}_{\mathbf{h}_{p+1}} = \mathbf{A}\hat{\mathbf{V}}^f_{\mathbf{h}_p}\mathbf{A}^H + \mathbf{R}$, $\bar{\mathbf{m}}_{\mathbf{h}_{p+1}} = \mathbf{A}\hat{\mathbf{m}}^f_{\mathbf{h}_p}$.

End For

• **Backward recursion:**

1) Initialization: Set $\bar{\mathbf{V}}^g_{\mathbf{h}_{P-1}} = \infty$ and $\bar{\mathbf{m}}^g_{\mathbf{h}_P - 1} = \mathbf{0}_{L_h \times 1}$.

2) For $p = P-1, \ldots, 0$

a) Set $\bar{\mathbf{V}}_{\mathbf{h}_p} = \bar{\mathbf{V}}^g_{\mathbf{h}_p}$, $\bar{\mathbf{m}}_{\mathbf{h}_p} = \bar{\mathbf{m}}^g_{\mathbf{h}_p}$.

b) Substitute $\{\bar{\mathbf{V}}_{\mathbf{h}_p}, \bar{\mathbf{m}}_{\mathbf{h}_p}\}$ into (22) and (4.25), to obtain $\{\hat{\mathbf{V}}_{\mathbf{h}_p}, \hat{\mathbf{m}}_{\mathbf{h}_p}\}$.

c) $\bar{\mathbf{V}}^g_{\mathbf{h}_{p-1}} = [\mathbf{A}^H(\hat{\mathbf{V}}_{\mathbf{h}_p} + \mathbf{R})^{-1}\mathbf{A}]^{-1}$, $\bar{\mathbf{m}}^g_{\mathbf{h}_{p-1}} = \mathbf{A}^{-1}\hat{\mathbf{m}}_{\mathbf{h}_p}$.

End For

• **Combination:**

For $p = 0, \ldots, P-1$

$$\begin{aligned} \hat{\mathbf{V}}_{\mathbf{h}_p} &= \left[\left(\hat{\mathbf{V}}^f_{\mathbf{h}_p}\right)^{-1} + \left(\bar{\mathbf{V}}^g_{\mathbf{h}_p}\right)^{-1}\right]^{-1}, \\ \hat{\mathbf{m}}_{\mathbf{h}_p} &= \hat{\mathbf{V}}_{\mathbf{h}_p}\left[\left(\hat{\mathbf{V}}^f_{\mathbf{h}_p}\right)^{-1}\hat{\mathbf{m}}^f_{\mathbf{h}_p} + \left(\bar{\mathbf{V}}^g_{\mathbf{h}_p}\right)^{-1}\bar{\mathbf{m}}^g_{\mathbf{h}_p}\right]. \end{aligned} \tag{4.27}$$

End For

where $E(\mathbf{w}_p)$ is calculated using (4.8), (4.9), and (4.10), by making the following replacements: $E(\mathbf{c}_p) = \hat{\mathbf{m}}_{\mathbf{c}_p}$, $E(\underline{\mathbf{t}}_p) = \mathbf{G}\hat{\mathbf{m}}_{\mathbf{t}_p}$, $v'_p = \frac{1}{N}\mathrm{tr}(\mathbf{G}\hat{\mathbf{V}}_{\mathbf{t}_\mathrm{p}}\mathbf{G}^{\mathrm{H}})$, and $[E(h_p^0), \ldots, E(h_p^{L_h-1})]^T = \hat{\mathbf{m}}_{\mathbf{h}_p}$. Although the CP is not inserted as that in conventional methods, the inter-subblock interference induced by FTN signalling and time-dispersive channels can be gradually cancelled out with the iteration by using more and more accurate estimates of the data symbols and the channel.

Note that in the considered turbo equalization system, the inner iteration between (4.28), (4.31) and (4.32) is successfully embedded into the outer iteration (i.e., the iteration between the equalizer and the decoder). In other words, in the Γth outer iteration, the decoder delivers the *extrinsic* mean vector and covariance matrix $\{\bar{\mathbf{m}}_{\mathbf{s}_p}, \bar{\mathbf{V}}_{\mathbf{s}_p}\}$ to the equalizer, which are fixed when the inner iteration is performed. After a certain number of inner iterations, the *extrinsic* information $p^e(\mathbf{s}_p) \propto \mathcal{N}_c(\mathbf{s}_p; \breve{\mathbf{m}}_{\mathbf{s}_p}, \breve{\mathbf{V}}_{\mathbf{s}_p})$ is fed back to the decoder to perform the $(\Gamma+1)$th outer iteration. The mean vector and the covariance matrix of $p^e(\mathbf{s}_p)$ are given by

$$\begin{aligned} \breve{\mathbf{V}}_{\mathbf{s}_p} &= [\hat{\mathbf{V}}_{\mathbf{s}_p}^{-1} - \bar{\mathbf{V}}_{\mathbf{s}_p}^{-1}]^{-1}, \\ \breve{\mathbf{m}}_{\mathbf{s}_p} &= \breve{\mathbf{V}}_{\mathbf{s}_p}[\hat{\mathbf{V}}_{\mathbf{s}_p}^{-1}\hat{\mathbf{m}}_{\mathbf{s}_p} - \bar{\mathbf{V}}_{\mathbf{s}_p}^{-1}\bar{\mathbf{m}}_{\mathbf{s}_p}]. \end{aligned} \tag{4.33}$$

We summarize the scheduling scheme of the proposed algorithm (we call this algorithm as JCED) in Algorithm 8.

By further exploiting the correlation between the elements in the frequency-domain CIRs, a refined JCED (R-JCED) algorithm is proposed in 4.3.2 to improve the accuracy of channel estimation. Moreover, in contrast to the Gaussian approximation of data symbols in JCED and R-JCED algorithms, we further employ the GAMP algorithm and incorporate it into the iteration between the channel estimation and decoder in 4.3.3, which leads to two algorithms, i.e., GAMP-JCED and R-GAMP-JCED.

4.3.2 R-JCED Based on the VB Framework

To further take account of the interaction between the elements in $\mathbf{c}_p$, it is found that the L_h-dimensional vector $\mathbf{c}_p$, constructed from L_h independent complex Gaussian

Algorithm 8 Joint channel estimation and decoding method (JCED).

• **Initialization:**

Set $\{\bar{\mathbf{m}}_{\mathbf{s}_p} = \mathbf{0}_{M\times 1}, \bar{\mathbf{V}}_{\mathbf{s}_p} = \mathbf{I}_M\}$, $\{\hat{\mathbf{m}}_{\mathbf{h}_p} = \mathbf{0}_{L_h\times 1}, \hat{\mathbf{V}}_{\mathbf{h}_p} = \infty\}$, and $\{\hat{\mathbf{z}}_p = \mathbf{y}_p, \hat{\mathbf{\Lambda}}_p = \mathbf{\Lambda}'_p\}$, for $p = 0, \ldots, P-1$.

• **Iteration:**

Outer Iteration:

For $\Gamma = 1, \ldots, \Gamma^{out}$

– **Inner Iteration:**

For $\gamma = 1, \ldots, \gamma^{in}$

Update $\{\hat{\mathbf{V}}_{\mathbf{h}_p}, \hat{\mathbf{m}}_{\mathbf{h}_p}\}$, $\{\hat{\mathbf{V}}_{\mathbf{s}_p}, \hat{\mathbf{m}}_{\mathbf{s}_p}\}$, and $\{\hat{\mathbf{z}}_p, \hat{\mathbf{\Lambda}}_p\}$, from (4.27), (4.30), and (4.32), respectively, for $p = 0, \ldots, P-1$.

End For

– Calculate the *extrinsic* information $\{\breve{\mathbf{V}}_{\mathbf{s}_p}, \breve{\mathbf{m}}_{\mathbf{s}_p}\}$ by (4.33), based on which and the coding constraints, the SISO decoder outputs the *extrinsic* information $\{\bar{\mathbf{s}}_p, \bar{\mathbf{V}}_{\mathbf{s}_p}\}$.

End For

random variables, has L_h degrees of freedom. Thus, by defining

$$\mathbf{c}_{p,0} = \sqrt{L_h}\mathbf{F}_{L_h}\mathbf{h}_p, \tag{4.34}$$

it can be further regarded as a specific group in $\mathbf{c}_p$, i.e., $\mathbf{c}_{p,0} = [c_p^0, c_p^U, \ldots, c_p^{(L_h-1)\times U}]^T$, where c_p^u is the uth element in $\mathbf{c}_p$, and $N = UL_h$ with U an integer (this can be realized properly by system design). In this way, the vector $\mathbf{c}_p$ is partitioned into U subvectors with length L_h as in [124], and the subvector at the index u is defined as

$$\mathbf{c}_{p,u} = [c_p^u, c_p^{U+u}, \ldots, c_p^{(L_h-1)\times U+u}]^T, \quad u = 0, 1, \ldots, U-1, \tag{4.35}$$

which can be rewritten as

$$\begin{aligned}\mathbf{c}_{p,u} &= \sqrt{L_h}\mathbf{F}_{L_h}\mathcal{D}([1, e^{-j2\pi u/N}, \ldots, e^{-j2\pi u(L_c-1)/N}])\mathbf{h}_p \\ &= \sqrt{L_h}\mathbf{F}_{L_h}\mathbf{D}_u\mathbf{h}_p \quad u = 0, 1, \ldots, U-1.\end{aligned} \tag{4.36}$$

Moreover, one can derive the following relationship between these subvectors

$$\begin{aligned}\mathbf{c}_{p,u-1} &= \mathbf{F}_{L_h}\mathcal{D}([1, e^{j2\pi/N}, \ldots, e^{j2\pi(L_h-1)/N}])\mathbf{F}_{L_h}^H\mathbf{c}_{p,u} \\ &\triangleq \mathbf{E}\mathbf{c}_{p,u}, \quad u = 1, \ldots, U-1.\end{aligned} \tag{4.37}$$

In a similar way, $\mathbf{y}_p$, $\underline{\mathbf{g}}$, $\mathbf{t}_p$ and $\mathbf{w}_p$ are all partitioned into U length-L_h subvectors, and the uth subvectors are respectively given by

$$\begin{aligned}\mathbf{y}_{p,u} &= [y_p^u, y_p^{U+u}, \ldots, y_p^{(L_h-1)\times U+u}]^T, \\ \underline{\mathbf{g}}_u &= [\underline{g}^u, \underline{g}^{U+u}, \ldots, \underline{g}^{(L_h-1)\times U+u}]^T, \\ \mathbf{t}_{p,u} &= [t_p^u, t_p^{U+u}, \ldots, t_p^{(L_h-1)\times U+u}]^T, \\ \mathbf{w}_{p,u} &= [w_p^u, w_p^{U+u}, \ldots, w_p^{(L_h-1)\times U+u}]^T,\end{aligned} \tag{4.38}$$

where y_p^u, $\underline{g}^u$, t_p^u and w_p^u are the uth elements of $\mathbf{y}_p$, $\underline{\mathbf{g}}$, $\mathbf{t}_p$ and $\mathbf{w}_p$, respectively.

With (4.35) and (4.38), the uth submodel of (4.4) is given by

$$\mathbf{y}_{p,u} = \mathbf{c}_{p,u} \odot \underline{\mathbf{g}}_u \odot \mathbf{t}_{p,u} + \mathbf{w}_{p,u}, \quad u = 0, 1, \ldots, U-1, \tag{4.39}$$

where

$$\begin{aligned}V(\mathbf{w}_{p,u}) &= \mathcal{D}([\Lambda_p^u, \Lambda_p^{U+u}, \ldots, \Lambda_p^{(L_h-1)\times U+u}]) \\ &\triangleq \mathbf{\Lambda}_{p,u}, \\ E(\mathbf{w}_{p,u}) &= [E_{\mathbf{w}_p}^u, E_{\mathbf{w}_p}^{U+u}, \ldots, E_{\mathbf{w}_p}^{(L_h-1)\times U+u}]^T,\end{aligned} \tag{4.40}$$

with Λ_p^u the uth diagonal element of $\mathbf{\Lambda}_p$, and $E_{\mathbf{w}_p}^u$ the uth element in $E(\mathbf{w}_p)$.

Comparing (4.4) with (4.39), it is observed that they have the same structure. Thereby, based on the submodel in (4.39), multiple independent estimates of the channel can be obtained within a subblock via VB framework as derived in Section III-B with some modifications. Define the posterior pdf of the channel calculated from the uth submodel by $p(\mathbf{h}_p; \hat{\mathbf{m}}_{\mathbf{h}_{p,u}}, \hat{\mathbf{V}}_{\mathbf{h}_{p,u}})$, the mean vector and covariance matrix are given by

$$\begin{aligned}\hat{\mathbf{V}}_{\mathbf{h}_{p,u}} &= \Big[L_h \mathbf{D}_u^H \mathbf{F}_{L_h}^H \Big(\mathcal{D}^H(\hat{\mathbf{m}}_{\underline{\mathbf{t}}_{p,u}})\hat{\mathbf{\Lambda}}_{p,u}^{-1}\mathcal{D}(\hat{\mathbf{m}}_{\underline{\mathbf{t}}_{p,u}}) \\ &\qquad +\mathcal{D}(\hat{\mathbf{V}}_{\underline{\mathbf{t}}_{p,u}})\hat{\mathbf{\Lambda}}_{p,u}^{-1}\Big)\mathbf{F}_{L_h}\mathbf{D}_u + \bar{\mathbf{V}}_{\mathbf{h}_{p,u}}^{-1}\Big]^{-1}, \\ \hat{\mathbf{m}}_{\mathbf{h}_{p,u}} &= \hat{\mathbf{V}}_{\mathbf{h}_{p,u}}\Big[\sqrt{L_h}\mathbf{D}_u^H\mathbf{F}_{L_h}^H\mathcal{D}^H(\hat{\mathbf{m}}_{\underline{\mathbf{t}}_{p,u}})\hat{\mathbf{\Lambda}}_{p,u}^{-1}\hat{\mathbf{z}}_{p,u} \\ &\qquad +\bar{\mathbf{V}}_{\mathbf{h}_{p,u}}^{-1}\bar{\mathbf{m}}_{\mathbf{h}_{p,u}}\Big],\end{aligned} \tag{4.41}$$

where $\hat{\mathbf{m}}_{\underline{\mathbf{t}}_{p,u}} = \mathcal{D}(\mathbf{g}_u)\hat{\mathbf{m}}_{\mathbf{t}_{p,u}}$, $\hat{\mathbf{V}}_{\underline{\mathbf{t}}_{p,u}} = \mathcal{D}(\mathbf{g}_u)\hat{\mathbf{V}}_{\mathbf{t}_{p,u}}\mathcal{D}^H(\mathbf{g}_u)$, $\hat{\mathbf{V}}_{\mathbf{t}_{p,u}}$, $\hat{\mathbf{m}}_{\mathbf{t}_{p,u}}$, $\hat{\mathbf{\Lambda}}_{p,u}$, and $\hat{\mathbf{z}}_{p,u}$ are obtained from the specific partition of $\hat{\mathbf{V}}_{\mathbf{t}_p}$, $\hat{\mathbf{m}}_{\mathbf{t}_p}$, $\hat{\mathbf{\Lambda}}_p$, and $\hat{\mathbf{z}}_p$, respectively. Then through a forward-backward recursion as described in Section III-B, the U independent channel estimates can be combined together. However, there exists complex matrix inversion in (4.41), which is computationally inefficient. To avoid this problem, we approximate $\hat{\mathbf{V}}_{\mathbf{h}_{p,u}}$ to be diagonal via similar methods as handling the inversion of $\hat{\mathbf{V}}_{\mathbf{h}_p}$ in (4.23). Accordingly, based on the relationship between $\mathbf{h}_{p,u}$ and $\mathbf{c}_{p,u}$ in (4.36), we rewrite (4.41) as

$$\begin{aligned}\hat{\mathbf{V}}_{\mathbf{c}_{p,u}} &= \Big[\mathcal{D}^H(\hat{\mathbf{m}}_{\mathbf{t}_{p,u}})\hat{\mathbf{\Lambda}}_{p,u}^{-1}\mathcal{D}(\hat{\mathbf{m}}_{\mathbf{t}_{p,u}}) + \mathcal{D}(\hat{\mathbf{V}}_{\mathbf{t}_{p,u}})\hat{\mathbf{\Lambda}}_{p,u}^{-1} + \bar{\mathbf{V}}_{\mathbf{c}_{p,u}}^{-1}\Big]^{-1} \\ \hat{\mathbf{m}}_{\mathbf{c}_{p,u}} &= \hat{\mathbf{V}}_{\mathbf{c}_{p,u}}\Big[\mathcal{D}^H(\hat{\mathbf{m}}_{\mathbf{t}_{p,u}})\hat{\mathbf{\Lambda}}_{p,u}^{-1}\hat{\mathbf{z}}_{p,u} + \bar{\mathbf{V}}_{\mathbf{c}_{p,u}}^{-1}\bar{\mathbf{m}}_{\mathbf{c}_{p,u}}\Big],\end{aligned} \tag{4.42}$$

and define the posterior pdf $q(\mathbf{c}_{p,u}) \propto \mathcal{N}_c(\mathbf{c}_{p,u}; \hat{\mathbf{m}}_{\mathbf{c}_{p,u}}, \hat{\mathbf{V}}_{\mathbf{c}_{p,u}})$, for $u = 0, \ldots, U-1$. The prior information $\{\bar{\mathbf{V}}_{\mathbf{c}_{p,u}}, \bar{\mathbf{m}}_{\mathbf{c}_{p,u}}\}$ is obtained from the $(u+1)$ *th* subvector through the forward recursion shown in Algorithm 9[e]. Apparently, the computational complexity is reduced compared with the conventional methods since it grows logarithmically with the length of the ISI taps of the unknown time-dispersive channels instead of the total number of ISI taps from both the FTN signalling and the channels.

[e]There is no need to perform backward recursion since we concern the update of the information corresponding to $\mathbf{c}_{p,0}$. Due to the page limit, the detailed derivation process of forward-backward recursion is not stated here.

Algorithm 9 Forward recursion based on (4.37).

1: • **Initialization:**

2: Set $\bar{\mathbf{V}}_{\mathbf{c}_{p,U-1}} = \infty$ and $\bar{\mathbf{m}}_{\mathbf{c}_{p,U-1}} = \mathbf{0}_{L_h \times 1}$.

3: • **Forward recursion:**

4: For $u = U-1, \ldots, 0$

5: - Substitute $\{\bar{\mathbf{V}}_{\mathbf{c}_{p,u}}, \bar{\mathbf{m}}_{\mathbf{c}_{p,u}}\}$ to (4.42), obtaining $\{\hat{\mathbf{V}}_{\mathbf{c}_{p,u}}, \hat{\mathbf{m}}_{\mathbf{c}_{p,u}}\}$.

6: - Calculate

$$\begin{aligned} \bar{\mathbf{V}}_{\mathbf{c}_{p,u-1}} &= \mathbf{E}\hat{\mathbf{V}}_{\mathbf{c}_{p,u}}\mathbf{E}^H \\ &\approx L_h^{-1}\mathrm{tr}(\hat{\mathbf{V}}_{\mathbf{c}_{\mathrm{p,u}}})\mathbf{I}_{\mathrm{L_h}\times 1}, \\ \bar{\mathbf{m}}_{\mathbf{c}_{p,u-1}} &= \mathbf{E}\hat{\mathbf{m}}_{\mathbf{c}_{p,u}}. \end{aligned} \tag{4.43}$$

7: End For

To further utilize the correlation of the channel coefficients between adjacent subblocks, we calculate $\check{\mathbf{V}}_{\mathbf{h}_p}$ and $\check{\mathbf{m}}_{\mathbf{h}_p}$ as

$$\begin{aligned} \check{\mathbf{V}}_{\mathbf{h}_p} &= L_h^{-1}\mathbf{F}_{L_h}^H \hat{\mathbf{V}}_{\mathbf{c}_{p,0}}\mathbf{F}_{L_h}, \\ \check{\mathbf{m}}_{\mathbf{h}_p} &= L_h^{-1/2}\mathbf{F}_{L_h}^H \hat{\mathbf{m}}_{\mathbf{c}_{p,0}}, \end{aligned} \tag{4.44}$$

which are then substituted into (4.19), resulting in

$$\begin{aligned} \hat{\mathbf{V}}_{\mathbf{h}_p} &= \left[L_h\mathbf{F}_{L_h}^H \hat{\mathbf{V}}_{\mathbf{c}_{p,0}}^{-1}\mathbf{F}_{L_h} + \bar{\mathbf{V}}_{\mathbf{h}_p}^{-1}\right]^{-1}, \\ \hat{\mathbf{m}}_{\mathbf{h}_p} &= \hat{\mathbf{V}}_{\mathbf{h}_p}\left[\sqrt{L_h}\mathbf{F}_{L_h}^H \hat{\mathbf{V}}_{\mathbf{c}_{p,0}}^{-1}\hat{\mathbf{m}}_{\mathbf{c}_{p,0}} + \bar{\mathbf{V}}_{\mathbf{h}_p}^{-1}\bar{\mathbf{m}}_{\mathbf{h}_p}\right]. \end{aligned} \tag{4.45}$$

By approximating $\bar{\mathbf{V}}_{\mathbf{h}_p}$ as $\bar{\beta}_{\mathbf{h}_p}\mathbf{I}_{L_h}$ and some averaging operations, Eq.(4.45) can be simplified as

$$\begin{aligned} \hat{\mathbf{V}}_{\mathbf{h}_p} &\approx \mathbf{F}_{L_h}^H\left(L_h\hat{\mathbf{V}}_{\mathbf{c}_{p,0}}^{-1} + \bar{\beta}_{\mathbf{h}_p}^{-1}\mathbf{I}_{L_h}\right)^{-1}\mathbf{F}_{L_h} \\ &\approx L_h^{-1}\mathrm{tr}\left[(\mathrm{L_h}\hat{\mathbf{V}}_{\mathbf{c}_{\mathrm{p},0}}^{-1} + \bar{\beta}_{\mathbf{h}_\mathrm{p}}^{-1}\mathbf{I}_{\mathrm{L_h}})^{-1}\right]\mathbf{I}_{\mathrm{L_h}} = \hat{\beta}_{\mathbf{h}_\mathrm{p}}\mathbf{I}_{\mathrm{L_h}}, \\ \hat{\mathbf{m}}_{\mathbf{h}_p} &\approx \hat{\mathbf{V}}_{\mathbf{h}_p}\left[\sqrt{L_h}\mathbf{F}_{L_h}^H \hat{\mathbf{V}}_{\mathbf{c}_{p,0}}^{-1}\hat{\mathbf{m}}_{\mathbf{c}_{p,0}} + \bar{\mathbf{V}}_{\mathbf{h}_p}^{-1}\bar{\mathbf{m}}_{\mathbf{h}_p}\right] \\ &\approx \sqrt{L_h}\mathbf{F}_{L_h}^H\left(L_h\hat{\mathbf{V}}_{\mathbf{c}_{p,0}}^{-1} + \bar{\beta}_{\mathbf{h}_p}^{-1}\mathbf{I}_{L_h}\right)^{-1}\hat{\mathbf{V}}_{\mathbf{c}_{p,0}}^{-1}\hat{\mathbf{m}}_{\mathbf{c}_{p,0}} + \frac{\hat{\beta}_{\mathbf{h}_p}}{\bar{\beta}_{\mathbf{h}_p}}\bar{\mathbf{m}}_{\mathbf{h}_p}. \end{aligned} \tag{4.46}$$

Finally, the forward-backward recursion is performed as in Algorithm 7.

In this way, the accuracy of the channel estimation can be improved, because the constraint in (4.37) is fully taken into account. The proposed R-JCED method is summarized in Algorithm 10.

4.3.3 GAMP-Aided JCED Algorithm (GAMP-JCED)

Note that all aforementioned methods treat the data symbols as Gaussian variables, since an exact discrete distribution over the signal constellation leads to intractable complexity. For example, it is observed that the computation for $q(\mathbf{s}_p)$ in (4.15) requires in general a complexity order of $\mathcal{O}(2^{JM})$ without Gaussian approximation. However, it has been shown that Gaussian assumption of symbols, which only utilizes the first and the second order moments of data symbols, leads to performance loss in frequency-domain equalization [75]. Moreover, the computational complexity is still a challenging task even though the symbols are modeled as Gaussian without the approximation in (4.17), where the variances of the *a prior* distribution of data symbols are replaced by their average, because a non-trivial matrix operation is involved in (4.29) with a computational complexity of order $\mathcal{O}(N^3)$ flops. Inspired by [120], the GAMP algorithm is employed, which can be easily embedded in the VB-framework-based method, to compute posterior pdfs of data symbols.

As shown that VB-framework-based algorithm can be interpreted as message passing on FG, GAMP is also a type of message passing method to approximate the SPA when the underlying FG is dense. When embedded in the VB-framework-based algorithm, the mean vector $\hat{\mathbf{m}}_{\mathbf{h}_p}$ and the covariance matrix $\hat{\mathbf{V}}_{\mathbf{h}_p}$ of the channel estimates calculated by (4.14) are used to construct the linear mixing matrix and to absorb the distortion due to the uncertainty of channel estimates into the noise, resulting in the following system model

$$\mathbf{y}_p = \mathbf{H}_p \mathbf{F}_N \underline{\mathbf{s}}_p + \tilde{\mathbf{w}}_p, \tag{4.47}$$

where $\mathbf{H}_p = \mathcal{D}(\hat{\mathbf{m}}_{\mathbf{c}_p})\mathbf{G}$ is a diagonal matrix, and $\underline{\mathbf{s}}_p = \mathbf{T}_3 \mathbf{s}_p$ is the symbol vector after padding $\mathbf{s}_p$ with $(N - M)$ zeros. The equivalent noise $\tilde{\mathbf{w}}_p = \mathbf{w}_p + \mathcal{D}(\mathbf{e}_{\mathbf{c}_p})\mathbf{G}\mathbf{F}_N \underline{\mathbf{s}}_p$, where $\mathbf{e}_{\mathbf{c}_p}$ is the zero-mean Gaussian random variable with covariance matrix $\hat{\mathbf{V}}_{\mathbf{c}_p}$.

Algorithm 10 Refined Joint Channel Estimation and Decoding (R-JCED) method.

- **Initialization:**
  - Set $\{\bar{\mathbf{m}}_{\mathbf{s}_p} = \mathbf{0}_{M\times 1}, \bar{\mathbf{V}}_{\mathbf{s}_p} = \mathbf{I}_M\}$, $\{\hat{\mathbf{m}}_{\mathbf{h}_p} = \mathbf{0}_{L_h\times 1}, \hat{\mathbf{V}}_{\mathbf{h}_p} = \infty\}$, and $\{\hat{\mathbf{z}}_p = \mathbf{y}_p, \hat{\mathbf{\Lambda}}_p = \mathbf{\Lambda}'_p\}$, for $p = 0, \ldots, P-1$.
- **Iteration:**
  - **Outer Iteration:**
  - For $\Gamma = 1, \ldots, \Gamma^{out}$
    - **Inner Iteration:**
      - For $\gamma = 1, \ldots, \gamma^{in}$
      - For $p = 0, \ldots, P-1$
        - 1) Update $\{\hat{\mathbf{V}}_{\mathbf{c}_{p,u}}, \hat{\mathbf{m}}_{\mathbf{c}_{p,u}}\}$ from (4.42) along with the forward recursion, for $u = 0, \ldots, U-1$.
        - 2) Update $\{\hat{\mathbf{V}}_{\mathbf{h}_p}, \hat{\mathbf{m}}_{\mathbf{h}_p}\}$ from (4.46) based on a forward-backward recursion process as described in Algorithm 7.
        - 3) Update $\{\hat{\mathbf{V}}_{\mathbf{s}_p}, \hat{\mathbf{m}}_{\mathbf{s}_p}\}$, and $\{\hat{\mathbf{z}}_p, \hat{\mathbf{\Lambda}}_p\}$, from (4.30), and (4.32), respectively.
      - End For
    - End For
    - Calculate the *extrinsic* information $\{\breve{\mathbf{V}}_{\mathbf{s}_p}, \breve{\mathbf{m}}_{\mathbf{s}_p}\}$ by (4.33), based on which and the coding constraints, the SISO decoder outputs the *extrinsic* information $\{\bar{\mathbf{m}}_{\mathbf{s}_p}, \bar{\mathbf{V}}_{\mathbf{s}_p}\}$.
  - End For

Therefore, the covariance matrix $\tilde{\mathbf{\Lambda}}_p$ of $\tilde{\mathbf{w}}_p$ can be calculated by

$$\begin{aligned}\tilde{\mathbf{\Lambda}}_p &= \hat{\mathbf{\Lambda}}_p + \mathbf{G}\mathcal{D}(\mathbf{F}_N\underline{\mathbf{s}}_p)\hat{\mathbf{V}}_{\mathbf{c}_p}\mathcal{D}^H(\mathbf{F}_N\underline{\mathbf{s}}_p)\mathbf{G}^H \\ &\approx \hat{\mathbf{\Lambda}}_p + \mathbf{G}\hat{\mathbf{V}}_{\mathbf{c}_p}\mathbf{G}^H. \end{aligned} \tag{4.48}$$

For brevity, the noiseless output vector is defined by $\boldsymbol{\xi}_p$, i.e., $\boldsymbol{\xi}_p = \mathbf{H}_p\mathbf{F}_N\underline{\mathbf{s}}_p$, and the linear transform matrix by $\mathbf{\Phi}_p = \mathbf{H}_p\mathbf{F}_N$. With a linear mixing matrix $\mathbf{H}_p\mathbf{F}_N$, the efficiency of the FFT can be maintained having a complexity growing logarithmically with the length of the subblock. Moreover, the GAMP relaxes the Gaussian prior constraint and incorporates output nonlinearities, which conjectures its superior performance to conventional MMSE-FDE algorithms [75].

Generally speaking, the following approximations, corresponding to the two *scalar estimation* functions denoted by $g_{out}(\cdot)$ and $g_{in}(\cdot)$, are the main ideas behind the derivation of the GAMP. First, the elements in $\boldsymbol{\xi}_p$, i.e., $\{\xi_n\}$ are assumed to be posterior independent and approximated as Gaussian based on the central limit theorem. The likelihood function of y_n given ξ_n is denoted by $p(y_n|\xi_n) = \mathcal{N}_c(\hat{z}_n;\xi_n,\tilde{\Lambda}_n)$ with $\xi_n = \sum_m[\mathbf{\Phi}_p]_{nm}\underline{s}_m$, $n = pM,\ldots,pM+N-1$, where $\underline{s}_n$, $\hat{z}_n$ and $\tilde{\Lambda}_n$ are the nth elements of $\underline{\mathbf{s}}_p$, $\hat{\mathbf{z}}_p$ and $\tilde{\mathbf{\Lambda}}_p$, respectively. In this way, the posterior pdf of ξ_n is approximated by

$$\begin{aligned} q(\xi_n|\mathbf{y}_p;\bar{m}_{\xi_n},\bar{V}_{\xi_n}) &= \frac{p(y_n|\xi_n)\mathcal{N}_c(\xi_n;\bar{m}_{\xi_n},\bar{V}_{\xi_n})}{\int_{\xi_n'} p(y_n|\xi_n')\mathcal{N}_c(\xi_n';\bar{m}_{\xi_n},\bar{V}_{\xi_n})} \\ &= \mathcal{N}_c(\xi_n;\hat{m}_{\xi_n},\hat{V}_{\xi_n}), \end{aligned} \tag{4.49}$$

with

$$\hat{V}_{\xi_n} = \frac{\tilde{\Lambda}_n\bar{V}_{\xi_n}}{\tilde{\Lambda}_n+\bar{V}_{\xi_n}}, \quad \hat{m}_{\xi_n} = \frac{\bar{V}_{\xi_n}\hat{z}_n + \hat{\Lambda}_n\bar{m}_{\xi_n}}{\tilde{\Lambda}_n+\bar{V}_{\xi_n}}, \tag{4.50}$$

where $\bar{V}_{\xi_n}$ and $\bar{m}_{\xi_n}$ are updated iteratively in the GAMP algorithm, the update rules of which are given later (see (4.57) and (4.58) in Algorithm 5). From [118], the output scalar function $g_{out}(\cdot)$ for the approximation of SPA to calculate the MMSE estimates is give by

$$g_{out}(\bar{m}_{\xi_n},y_n,\bar{V}_{\xi_n}) = \frac{1}{\bar{V}_{\xi_n}}[E(\xi_n|\mathbf{y}_p;\bar{m}_{\xi_n},\bar{V}_{\xi_n}) - \bar{m}_{\xi_n}]$$

$$= \frac{1}{\bar{V}_{\xi_n}} \left(\frac{\bar{V}_{\xi_n} \hat{z}_n + \tilde{\Lambda}_n \bar{m}_{\xi_n}}{\tilde{\Lambda}_n + \bar{V}_{\xi_n}} - \bar{m}_{\xi_n} \right)$$
$$= \frac{\hat{z}_n - \bar{m}_{\xi_n}}{\tilde{\Lambda}_n + \bar{V}_{\xi_n}}, \tag{4.51}$$

and its negative partial derivative of $g_{out}(\bar{m}_{\xi_n}, y_n, \bar{V}_{\xi_n})$ is

$$-\frac{\partial}{\partial \bar{m}_{\xi_n}} g_{out}(\bar{m}_{\xi_n}, y_n, \bar{V}_{\xi_n}) = \frac{1}{\tilde{\Lambda}_n + \bar{V}_{\xi_n}}. \tag{4.52}$$

Second, the posterior pdfs of the elements $\{s_n\}$, $n = pM, \ldots, pM + M - 1$, in the symbol vector $\underline{\mathbf{s}}_p$ are assumed to be independent, and the true posterior distribution $p(s_n|\mathbf{y})$ is approximated as

$$q(s_n = a_i|\mathbf{y}_p; \check{m}_{\eta_n}, \check{V}_{\eta_n})$$
$$= \frac{p(s_n = a_i)\mathcal{N}_c(s_n = a_i|\check{m}_{\eta_n}, \check{V}_{\eta_n})}{\sum_{a_i \in \mathcal{X}} p(s_n = a_i)\mathcal{N}_c(s_n = a_i|\check{m}_{\eta_n}, \check{V}_{\eta_n})}, \tag{4.53}$$

for $n = pM, \ldots, pM + M - 1$, where $\check{V}_{\eta_n}$ and $\check{m}_{\eta_n}$ are updated iteratively in the GAMP algorithm, the update rules of which are given later (see (4.60) and (4.61) in Algorithm 5). The input scalar function is related to the posterior mean of $\underline{s}_n$ as follows:

$$g_{in}(\check{m}_{\eta_n}, \check{V}_{\eta_n}) = E(s_n|\mathbf{y}_p; \check{m}_{\eta_n}, \check{V}_{\eta_n})$$
$$= \sum_{a_i \in \mathcal{X}} a_i q(s_n = a_i|\mathbf{y}_p; \check{m}_{\eta_n}, \check{V}_{\eta_n}), \tag{4.54}$$

for $n = pM, \ldots, pM+M-1$, and $g_{in}(\check{m}_{\eta_n}, \check{V}_{\eta_n}) = 0$ for $n = pM+M, \ldots, pM+N-1$, where $E\{\cdot\}$ is the expectation operation. Similarly, the derivative of $g_{in}(\check{m}_{\eta_n}, \check{V}_{\eta_n})$ satisfies

$$\check{V}_{\eta_n} \frac{\partial}{\partial \check{m}_{\eta_n}} g_{in}(\check{m}_{\eta_n}, \check{V}_{\eta_n}) = V(s_n|\mathbf{y}_p; \check{m}_{\eta_n}, \check{V}_{\eta_n})$$
$$= \sum_{a_i \in \mathcal{X}} |a_i - g_{in}(\check{m}_{\eta_n}, \check{V}_{\eta_n})|^2$$
$$\times q(s_n|\mathbf{y}_p; \check{m}_{\eta_n}, \check{V}_{\eta_n}), \tag{4.55}$$

for $n = pM, \ldots, pM+M-1$, and $\check{V}_{\eta_n} \frac{\partial}{\partial \check{m}_{\eta_n}} g_{in}(\check{m}_{\eta_n}, \check{V}_{\eta_n}) = 0$, for $n = pM+M, \ldots, pM+N-1$.

Algorithm 11 The symbol detection based on the GAMP algorithm.

• **Initialization:**

For $p = 0, \ldots, P-1$

Set $\check{m}_{\eta_n} = 0$, $\check{V}_{\eta_n}^{-1} = 0$, and $\hat{m}_{\zeta_n} = 0$, for $n = pM, \ldots, pM + N - 1$.

End For

• **Iteration:**

For $p = 0, \ldots, P-1$

For $n = pM, \ldots, pM + N - 1$

1) Calculate the posterior mean and posterior variance for $\underline{s}_n$ by

$$\begin{aligned} \hat{m}_{\underline{s}_n} &\triangleq g_{in}(\check{m}_{\eta_n}, \check{V}_{\eta_n}), \\ \hat{V}_{\underline{s}_n} &\triangleq \check{V}_{\eta_n} \frac{\partial}{\partial \check{m}_{\eta_n}} g_{in}(\check{m}_{\eta_n}, \check{V}_{\eta_n}), \end{aligned} \tag{4.56}$$

for $n = pM, \ldots, pM + M - 1$, and $\hat{m}_{\underline{s}_n} = 0$, $\hat{V}_{\underline{s}_n} = 0$, for $n = pM + M, \ldots, pM + N - 1$.

2) Calculate $\bar{V}_{\xi_n}$ and $\bar{m}_{\xi_n}$ by

$$\bar{V}_{\xi_n} = \sum_m |[\mathbf{\Phi}_p]_{nm}|^2 \hat{V}_{\underline{s}_m}, \tag{4.57}$$

$$\bar{m}_{\xi_n} = \sum_m [\mathbf{\Phi}_p]_{nm} \hat{m}_{\underline{s}_m} - \bar{V}_{\xi_n} \hat{m}_{\zeta_n}. \tag{4.58}$$

3) Calculate

$$\begin{aligned} \hat{m}_{\zeta_n} &\triangleq g_{out}(\bar{m}_{\xi_n}, y_n, \bar{V}_{\xi_n}), \\ \hat{V}_{\zeta_n} &\triangleq -g'_{out}(\bar{m}_{\xi_n}, y_n, \bar{V}_{\xi_n}). \end{aligned} \tag{4.59}$$

4) Calculate $\check{V}_{\eta_n}$ and $\check{m}_{\eta_n}$ by

$$\check{V}_{\eta_n} = (\sum_m |[\mathbf{\Phi}_p]_{mn}|^2 \hat{V}_{\zeta_m})^{-1}, \tag{4.60}$$

$$\check{m}_{\eta_n} = \hat{m}_{\underline{s}_n} + \check{V}_{\eta_n} \sum_m [\mathbf{\Phi}_p]^*_{mn} \hat{m}_{\zeta_m}. \tag{4.61}$$

End For

End For

Given the input and output scalar functions, the iterative GAMP algorithm for updating the estimates of ξ_n and s_n is summarized in Algorithm 11.

Note that the complexity of the GAMP algorithm is $\mathcal{O}(N^2)$, dominated by the matrix-vector multiplications in (4.57), (4.58), (4.60) and (4.61). However, the overall complexity can be reduced to $\mathcal{O}(N \log N)$ thanks to the FFT. Specifically, by using the approximation $\mathbf{F}_N \mathcal{D}(\hat{\mathbf{V}}_{\underline{\mathbf{s}}_p}) \mathbf{F}_N^H \approx (\frac{1}{N} \mathbf{1}_{N\times 1}^T \hat{\mathbf{V}}_{\underline{\mathbf{s}}_p}) \mathbf{I}_N$, Eq. (4.57) can be rewritten as

$$\begin{aligned} \bar{\mathbf{V}}_{\boldsymbol{\xi}_p} &= |\mathbf{H}_p \mathbf{F}_N|^2 \hat{\mathbf{V}}_{\underline{\mathbf{s}}_p} = \mathcal{D}(\mathbf{H}_p \mathbf{F}_N \mathcal{D}(\hat{\mathbf{V}}_{\underline{\mathbf{s}}_p}) \mathbf{F}_N^H \mathbf{H}_p^H) \mathbf{1}_{N\times 1} \\ &\approx (\frac{1}{N} \mathbf{1}_{N\times 1}^T \hat{\mathbf{V}}_{\underline{\mathbf{s}}_p}) \mathbf{H}_p \mathbf{H}_p^H \mathbf{1}_{N\times 1}, \end{aligned} \tag{4.62}$$

where the squared magnitude is taken componentwisely, $\hat{\mathbf{V}}_{\underline{\mathbf{s}}_p} = [\hat{V}_{s_{pM}}, \ldots, \hat{V}_{s_{pM+M-1}}, \mathbf{0}_{1\times(N-M)}]^T$ and $\bar{\mathbf{V}}_{\boldsymbol{\xi}_p} = [\bar{V}_{\xi_{pM}}, \ldots, \bar{V}_{\xi_{pM+N-1}}]^T$. Also, we have

$$\bar{\mathbf{m}}_{\boldsymbol{\xi}_p} = \mathbf{H}_p \mathbf{F}_N \hat{\mathbf{m}}_{\underline{\mathbf{s}}_p} - \bar{\mathbf{V}}_{\boldsymbol{\xi}_p} \odot \hat{\mathbf{m}}_{\boldsymbol{\zeta}_p}, \tag{4.63}$$

where $\bar{\mathbf{m}}_{\boldsymbol{\xi}_p} \triangleq [\bar{m}_{\xi_{pM}}, \ldots, \bar{m}_{\xi_{pM+N-1}}]^T$, $\hat{\mathbf{m}}_{\underline{\mathbf{s}}_p} \triangleq [\hat{m}_{s_{pM}}, \ldots, \hat{m}_{s_{pM+M-1}}, \mathbf{0}_{1\times(N-M)}]^T$, and $\hat{\mathbf{m}}_{\boldsymbol{\zeta}_p} \triangleq [\hat{m}_{\zeta_{pM}}, \ldots, \hat{m}_{\zeta_{pM+N-1}}]^T$. Similarly, by using the approximation $\mathbf{F}_N^H \mathbf{H}_p^H \mathcal{D}(\hat{\mathbf{V}}_{\boldsymbol{\zeta}_p}) \mathbf{H}_p \mathbf{F}_H \approx \mathrm{tr}(\mathbf{H}_{\mathrm{p}}^{\mathrm{H}} \mathcal{D}(\hat{\mathbf{V}}_{\boldsymbol{\zeta}_{\mathrm{p}}}) \mathbf{H}_{\mathrm{p}})$, (4.60) can also be represented in a vector form, i.e.,

$$\begin{aligned} \mathbf{1}_{N\times 1} ./ \check{\mathbf{V}}_{\boldsymbol{\eta}_p} &= |\mathbf{F}_N^H \mathbf{H}_p^H|^2 \hat{\mathbf{V}}_{\boldsymbol{\zeta}_p} \\ &= \mathcal{D}(\mathbf{F}_N^H \mathbf{H}_p^H \mathcal{D}(\hat{\mathbf{V}}_{\boldsymbol{\zeta}_p}) \mathbf{H}_p \mathbf{F}_H) \mathbf{1}_{N\times 1} \\ &\approx (\frac{1}{N} \mathbf{1}_{N\times 1}^H \mathbf{H}_p^H \mathbf{H}_p \hat{\mathbf{V}}_{\boldsymbol{\zeta}_p}) \mathbf{1}_{N\times 1} \triangleq \beta_{\boldsymbol{\eta}_p}^{-1} \mathbf{1}_{N\times 1}, \end{aligned} \tag{4.64}$$

where $\check{\mathbf{V}}_{\boldsymbol{\eta}_p} \triangleq [\check{V}_{\eta_{pM}}, \ldots, \check{V}_{\eta_{pM+N-1}}]^T$ and $\hat{\mathbf{V}}_{\boldsymbol{\zeta}_p} \triangleq [\hat{V}_{\zeta_{pM}}, \ldots, \hat{V}_{\zeta_{pM+N-1}}]^T$. Hence, (4.61) is reduced to the following equation

$$\check{\mathbf{m}}_{\boldsymbol{\eta}_p} = \hat{\mathbf{m}}_{\underline{\mathbf{s}}_p} + \beta_{\boldsymbol{\eta}_p} \mathbf{F}_N^H \mathbf{H}_p^H \hat{\mathbf{m}}_{\boldsymbol{\zeta}_p}, \tag{4.65}$$

where $\check{\mathbf{m}}_{\boldsymbol{\eta}_p} \triangleq [\check{m}_{\eta_{pM}}, \ldots, \check{m}_{\eta_{pM+N-1}}]^T$ and $\hat{\mathbf{m}}_{\boldsymbol{\zeta}_p} \triangleq [\hat{m}_{\zeta_{pM}}, \ldots, \hat{m}_{\zeta_{pM+N-1}}]^T$. It is observed that within the iteration of GAMP, some of the matrix-vector multiplications can be efficiently performed by FFT. The calculational complexity of other matrix-vector multiplications is low, because diagonal matrices are involved.

From the derivation of the GAMP, it is easily analyzed that the posterior mean and variance of the data sequence $\underline{\mathbf{s}}_p$ are approximately calculated as $\hat{\mathbf{m}}_{\underline{\mathbf{s}}_p}$ and $\hat{\mathbf{V}}_{\underline{\mathbf{s}}_p}$,

respectively. Thereby, the posterior pdf of $\mathbf{t}_p$ can be approximated to be Gaussian, i.e., $q(\mathbf{t}_p) \propto \mathcal{N}_c(\mathbf{t}_p; \hat{\mathbf{m}}_{\mathbf{t}_p}, \hat{\mathbf{V}}_{\mathbf{t}_p})$, with

$$\begin{aligned} \hat{\mathbf{m}}_{\mathbf{t}_p} &= \mathbf{F}_N \hat{\mathbf{m}}_{\mathbf{s}_p}, \\ \hat{\mathbf{V}}_{\mathbf{t}_p} &= \mathbf{F}_N \hat{\mathbf{V}}_{\mathbf{s}_p} \mathbf{F}_N^H \approx \frac{1}{M} \mathrm{tr}(\hat{\mathbf{V}}_{\mathbf{s}_\mathrm{p}}) \mathbf{I}_\mathrm{N}. \end{aligned} \tag{4.66}$$

Given (4.66), the channel estimation can be performed as the same in JCED and R-JCED, which are referred to as GAMP-JCED and R-GAMP-JCED, respectively. Most important, the GAMP method can be easily incorporated into the iteration between the SISO equalization and the SISO decoder. According to the turbo principle, $\check{V}_{\eta_n}$ and $\check{m}_{\eta_n}$ are separated from the *a priori* of the data symbols s_n, $n = pM, \ldots, pM + M - 1$, and thus they can be regarded as the extrinsic variance and the extrinsic mean, which are collected to vectors as

$$\begin{aligned} \check{\mathbf{V}}_{\mathbf{s}_p} &= \mathcal{D}([\check{V}_{\eta_{pM}}, \ldots, \check{V}_{\eta_{pM+M-1}}]^T), \\ \check{\mathbf{m}}_{\mathbf{s}_p} &= [\check{m}_{\eta_{pM}}, \ldots, \check{m}_{\eta_{pM+M-1}}]^T. \end{aligned} \tag{4.67}$$

The GAMP-JCED and the R-GAMP-JCED methods are summarized in Algorithm 12.

4.3.4 Complexity Analysis

In Table 1, a brief complexity analysis of the proposed JCED, R-JCED, GAMP-JCED, and R-GAMP-JCED algorithms is given compared with the existing methods [17,57–59,110,124,125]. All of these methods involve the calculations of $\mathbf{y}_p$ requiring P N-point FFTs operations with complexity $PN \log N$. Since the standard BCJR decoding is performed for all methods, only the computational complexity for one inner iteration within a subblock is analyzed in the following. In JCED, one N-point IFFT is needed for the calculation of (4.25), and one N-point FFT for (4.28). In R-JCED, $2(U-1)$ L_h-point FFTs are needed for the calculation of (4.43), one L_h-point IFFT for (4.46), and one N-point FFT for (4.28). The GAMP-JCED and the JCED differ in the equalization part. In JCED, one N-point IFFT is required for the calculation of (4.30), one N-point FFT for (4.31), and two N-point (I)FFTs for (4.10). In GAMP-JCED, one N-point FFT for (4.63), N-point FFT for (4.65), N-point FFT for (4.66), and two N-point (I)FFTs for (4.10). The CP-based [17,57,59]

Algorithm 12 The GAMP-JCED and R-GAMP-JCED algorithms.

- GAMP-JCED: The same as JCED except for the two modifications:
  - Inner iteration:
    Update $\{\hat{\mathbf{m}}_{\mathbf{s}_p}, \hat{\mathbf{V}}_{\mathbf{s}_p}\}$ from (4.56) instead of (4.30).
  - Outer iteration:
    Calculate the *extrinsic* information $\{\check{\mathbf{V}}_{\mathbf{s}_p}, \check{\mathbf{m}}_{\mathbf{s}_p}\}$ by (4.67) instead of (4.33).
- R-GAMP-JCED: The same as R-JCED except for the two modifications:
  - Inner iteration:
    Update $\{\hat{\mathbf{m}}_{\mathbf{s}_p}, \hat{\mathbf{V}}_{\mathbf{s}_p}\}$ from (4.56) instead of (4.30), then parse them into U subvectors with the uth subvector denoted by $\hat{\mathbf{m}}_{\mathbf{t}_{p,u}}$ and $\hat{\mathbf{V}}_{\mathbf{t}_{p,u}}$, respectively.
  - Outer iteration:
    Calculate the *extrinsic* information $\{\check{\mathbf{V}}_{\mathbf{s}_p}, \check{\mathbf{m}}_{\mathbf{s}_p}\}$ by (4.67) instead of (4.33).

Table 4.1: Comparisons of computational complexity

Algorithm	Complexity of the equalizer	Complexity of the channel estimator
JCED	$4N \log N$	$2N \log N$
GAMP-JCED	$5N \log N$	$2N \log N$
R-JCED	$4N \log N$	$(2U-1)L_h \log L_h + N \log N$
R-GAMP-JCED	$5N \log N$	$(2U-1)L_h \log L_h + N \log N$
VB-approx	$4N \log N$	$2N \log N$
EM-based	$4N \log N$	$(2U-1)L \log L + N \log N$
LS-based	$4N \log N$	$(2U-1)L \log L + N \log N$
CP-based (CP-based-w)	$2N \log N$	-
Overlap-based	$2N \log N$	-
IB-DFE	$2N \log N$	-

and overlap-based [58] algorithms requires $2N \log N$ complexity for equalization with known CSI. In EM-based [110] and LS-based [124] algorithms, the complexity of the channel estimation is $(2U-1)L \log L + N \log N$ and that of detection is $4N \log N$.

For VB-approx [125], the complexity of the channel estimation and the detection are $2N \log N$ and $4N \log N$, respectively.

4.4 Simulation Results

In this simulation, a rate-1/2 LDPC code[f] with block size 4896 and a QPSK signal with Gray mapping is considered. The root raised cosine filter with a roll-off factor $\alpha = 0.5$ is implemented for both the transmitting and the receiving filters. The packing ratio in the range from $\tau = 0.9$ to 0.6 is used for FTN signalling. For the time-varying frequency-selective fading channels, the carrier frequency is 2 GHz with 2.5 μs symbol period and mobile velocity is 110 km/h, i.e., the Doppler spread is 200Hz and the normalized fading rate is 0.0005. The total number of taps of the time-varying frequency-selective fading channels is $L_h = 64$, if not specified otherwise. The power delay profile is chosen as $\boldsymbol{\psi} = [1, e^{-0.1}, \ldots, e^{-0.1(L_h-1)}]^T$. A periodically extended 25th root length-139 Chu sequence [127] is employed as the training sequence and superimposed onto the data symbols, with power ratio of pilots to data symbols $\varsigma = 1/10$. The corresponding power loss due to the superimposed training sequence is about 0.4 dB, which has been considered in E_b/N_0. At the receiver, the value of ν is set to 9, unless otherwise specified. The number of the outer and inner iterations are 20 and 5, respectively. The number of iterations for LDPC decoding is 50.

The BER performance of the proposed JCED algorithm is evaluated and compared with the sate-of-the-art methods, namely, CP-based MMSE-FDE [17] (denoted by "CP-based"), CP-based MMSE-FDE with noise whitening [57] (denoted by "CP-based-w"), overlap-based MMSE-FDE [58] (denoted by "Overlap-based"), IB-DFE [59] (denoted by "IB-DFE") and the counterpart system with Nyquist rate (denoted by "Nyquist") in Fig. 4.2. The impacts of different packing ratios in FTN signalling are also studied. Since CIRs are assumed to be perfectly known in [17, 57, 58] and [59], all mentioned algorithms are first evaluated with known

[f]The variable and check node degree distributions are $X_1 = 0.0005 + 0.2285X + 0.0857X^2 + 0.6857X^7$ and $X_2 = 0.2290X^7 + 0.7710X^8$, respectively. The parameter definitions follow [72].

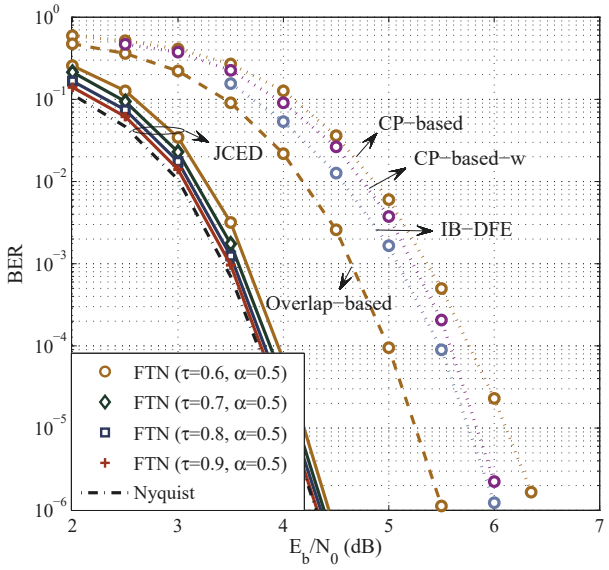

Figure 4.2: BER performance of different algorithms with known CIRs.

CIRs for a fair comparison. The "known CIRs" means that the CIRs in the central point of a subblock are assumed to be known and selected as the one that this subblock undergoes in the receiver. Note that though we assume that the CIRs are fixed within a subblock duration, the time-varying channels are generated by Jake's simulator in all simulations. With known CIRs, the proposed R-JCED algorithm is equivalent to the JCED method. The length of the subblock is set as $M = 129$ for the Nyquist case (i.e., $P = 19$, $Mf_DT_s = 0.065$), while $M = 175$ for other FTN cases (i.e., $P = 14$, $\tau Mf_DT_s = 0.088$)[g]. Due to the white noise assumption in the CP-based method in [17], the performance loss compared with the proposed JCED algorithm is about 2 dB at BER=10^{-5}. This is because that even though the colored noise can be recognized to be approximately independent in frequency domain, the amplitude of the autocorrelation coefficients may fluctuate greatly. Therefore the approximation error in the CP-based method will be large. With noise whitening, the CP-based-w method in [57] is superior to the CP-based method. The

[g]The Nyquist case is the lower bound of the FTN signalling case, since the length of the subblock is shorter than that selected for the latter one. The design of the subblock length is flexible, and it is better to ensure that blocks are independent, since decoding can be performed without latency, however, zeros-padding is needed between different blocks to eliminate the ISI effect. The selection of M in this paper is just for convenience.

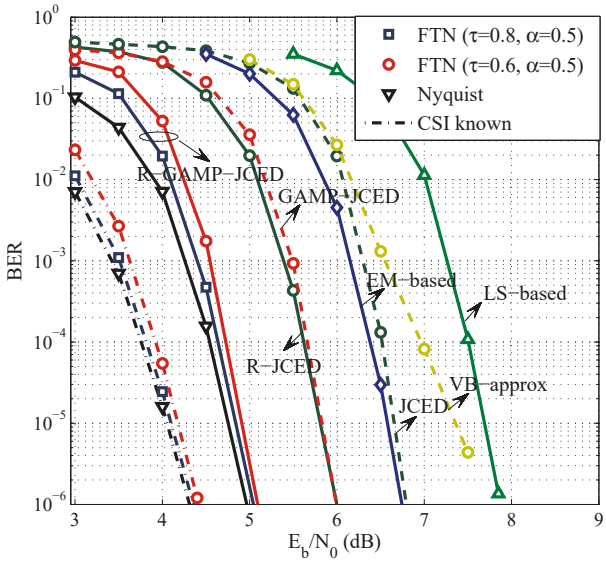

Figure 4.3: BER performance of the proposed algorithms with unknown CIR.

IB-DFE [59] delivers better BER performance compared with CP-based and CP-based-w algorithms due to the iterative feedback process. However, the CP used in the CP-based, CP-based-w and IB-DFE methods introduces about 30% overhead, which results in significant power and spectral efficiency loss. For the overlap-based method, the parameters p and q defined in [58] are set to 45, which means the length of overlap between adjacent subblocks is 90. Although the overlap is longer than that used in the proposed algorithms, BER performance is still inferior to that of the proposed ones, because the ISI introduced from the fading channels and FTN signalling is not sufficiently suppressed. It is also observed that performance loss of the FTN systems compared with the Nyquist counterpart is negligible even for $\tau = 0.6$ at BER=10^{-6}, which demonstrates the effectiveness of the approximations of the colored noise in (4.7) and the interference from adjacent subblocks in (4.8) due to the absence of CP.

In Fig. 4.3, we plot the BER performance of the proposed JCED, R-JCED, GAMP-JCED and R-GAMP-JCED algorithms for both FTN signalling and Nyquist signalling over unknown time-varying frequency-selective fading channels. Two packing ratios for the FTN signalling are considered, i.e., $\tau = 0.6$ and 0.8. The BER curves of the EM-based (denoted by "EM-based") [110], LS-based (denoted

by "LS-based") [124], VB-based (denoted by "VB-approx") [125], and the known CSI cases are also plotted for comparison. It is seen that GAMP-aided methods outperform the corresponding JCED methods, e.g., about 0.8 dB performance gain can be obtained by comparing GAMP-JCED (R-GAMP-JCED) with JCED (R-JCED). The superior performance of GAMP-aided methods is because data symbols are treated as discrete random variables in GAMP-JCED and R-GAMP-JCED, instead of being approximated as Gaussian distributions in JCED and R-JCED algorithms. We also observe that the refined channel estimate in R-JCED and R-GAMP-JCED leads to superior performance relative to the corresponding JCED and GAMP-JCED algorithms. For the cases $\tau = 0.8$ and $\tau = 0.6$, performance of the R-GAMP-JCED is very close to that of the Nyquist signalling with unknown CSI. That means that up to 67% higher transmission rate can be obtained with $\alpha = 0.5$ by employing the proposed R-GAMP-JCED algorithm in FTN system, with negligible BER performance loss in terms of E_b/N_0. Since the EM-based and LS-based methods in [110] and [124] were originally proposed for Nyquist signalling, we have extended them directly to FTN signalling via combining the ISI induced by both FTN signalling and the unknown frequency-selective channels together. Compared to the proposed R-JCED algorithm, about 0.7 dB performance loss can be observed for the EM-based method. This is mainly due to the fact that EM-based method ignored the knowledge of the known ISI induced by FTN signalling. On the contrary, the proposed R-JCED algorithm is able to fully exploit the structure of FTN signalling and only estimate the CIRs of the unknown frequency-selective channels. The LS-based algorithm suffers extra performance loss due to the separate process of the channel estimation and data detection compared with the EM-based counterpart. We also observe that the diagonal approximation of $\check{\mathbf{V}}_{\mathbf{h}_p}$ employed in [125] delivers poor performance in high E_b/N_0 due to the fluctuation of the amplitude of the entries in $\hat{\mathbf{m}}_{\underline{\mathbf{t}}_p}$.

The mean-squared errors (MSEs) of the channel estimation with different algorithms are illustrated in Fig. 4.4. It can be seen that the performance gap between the proposed R-GAMP-JCED algorithm for FTN signalling and the counterpart Nyquist system is marginal especially when E_b/N_0 is greater than a certain value.

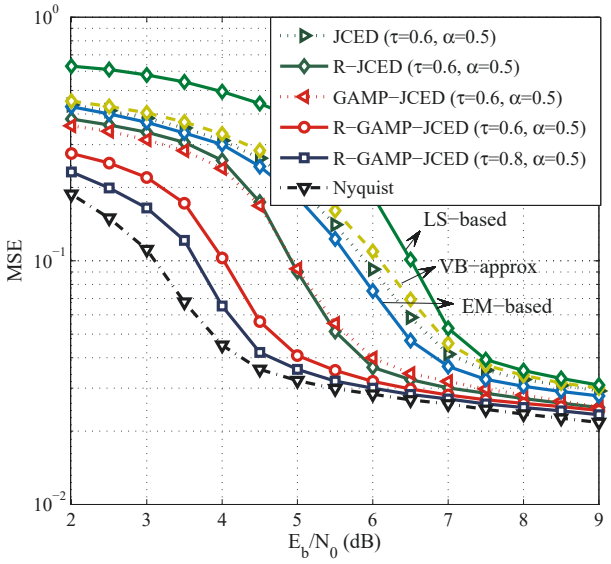

Figure 4.4: MSEs of channel estimation of different algorithms with unknown CIR.

The accuracy of channel estimation is improved with larger packing ratio at lower E_b/N_0. By exploiting the correlations between the elements in $\mathbf{c}_p$, the algorithms with refined channel estimation, namely, R-JCED and R-GAMP-JCED, significantly outperform the ones that ignore this information. Moreover, the GAMP-aided methods outperform the JCED algorithms, which is also reasonable since the performance of channel estimation can be improved by more accurate data symbol detection via GAMP in turbo receivers. Note that the MSE error-floor is caused by the time-varying channel state information that cannot be perfectly estimated.

Considering that the amount of ISI for the FTN signalling highly depends on the roll-off factor, the BER performance of the proposed R-JCED and R-GAMP-JCED is evaluted with roll-off factor $\alpha = 0.1$ to 0.5 in Fig. 4.5. It is observed that, for a fixed packing ratio τ, the BER performance degrades while reducing the roll-off factor. Note that this BER loss is especially obvious for low-τ scenarios. However, the spectral efficiency gain is obtained for lower packing ratio and roll-off factor. For example, compared with $\{\tau = 0.8, \alpha = 0.5\}$ ($\eta = 0.83$ bits/s/Hz), about 80% spectral efficiency gain is obtained by using $\{\tau = 0.6, \alpha = 0.1\}$ (η =1.52 bits/s/Hz), with about 0.3 dB E_b/N_0 loss at BER= 10^{-6} when the R-GAMP-JCED is employed.

In Fig. 4.6, the impact of different training sequences employed is compared,

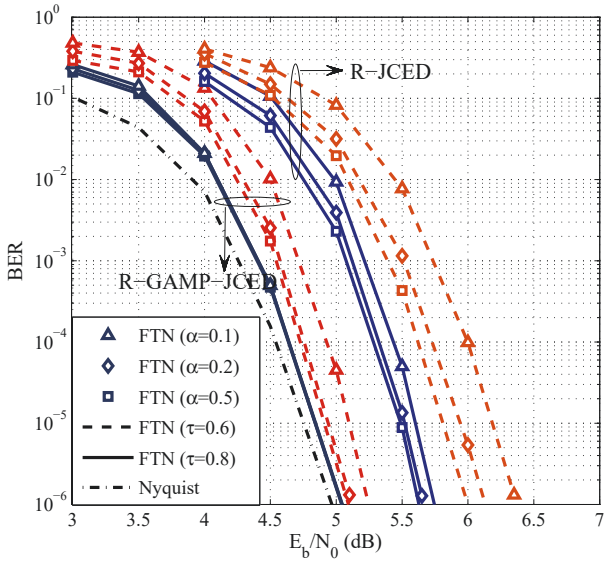

Figure 4.5: BER performance of the proposed R-JCED and R-GAMP-JCED algorithms with different roll-off factors.

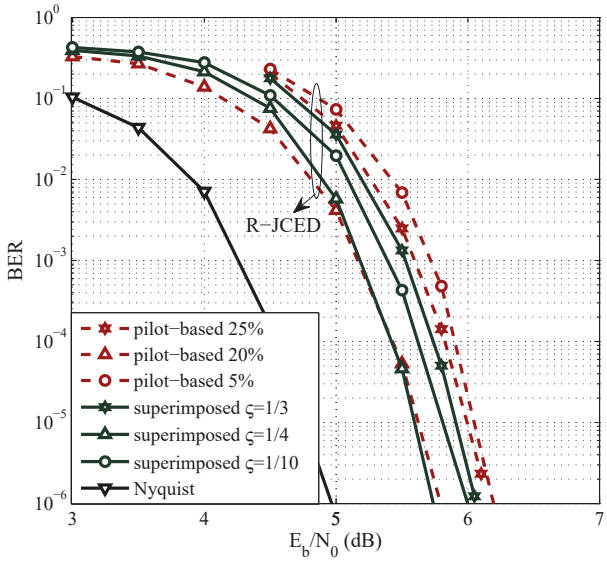

Figure 4.6: BER performance of the proposed R-JCED algorithm with different training sequences.

namely, the traditional pilot insertion (denoted by "pilot-based") and the superimposed training (denoted by "superimposed"). It is observed that, for both methods, increasing the ratio of pilots from 5% to 20% or increasing the power ratios from $\varsigma = 1/10$ to $\varsigma = 1/4$ lead to improved BER performance, owning to the more

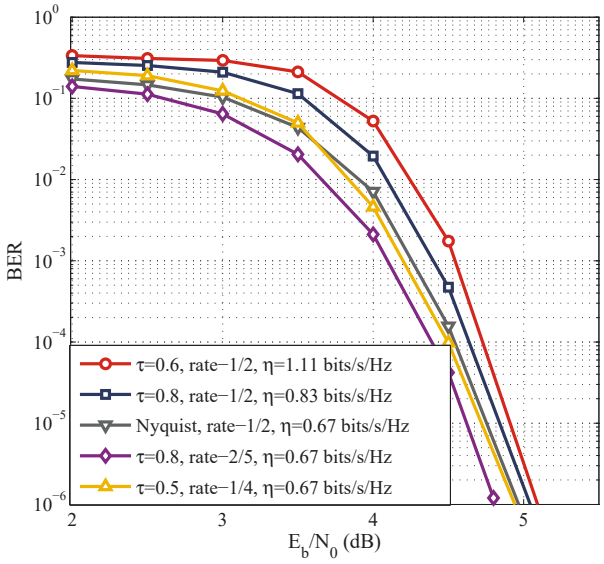

Figure 4.7: BER performance of the proposed R-GAMP-JCED algorithm with different τ and code rate.

accurate channel estimation obtained[h]. However, further increasing the ratio of pilots to 25% or the power ratio to $\varsigma = 1/3$ lead to performance loss, since the improvement in channel estimation becomes marginal and the loss in E_b/N_0 due to the training sequence dominates the BER performance. Moreover, we observe that the pilot-based method with 20% pilot ratio performs very close to the superimposed training with $\varsigma = 1/4$. However, the former expends the signal bandwidth by 20%, which obviously compromises the advantage of FTN signalling in spectral efficiency.

In Fig. 4.7, the BER performance of the proposed R-GAMP-JCED algorithm with different values of τ and code rates is evaluated. The scenario of Nyquist signalling is also plotted for comparison. It is observed that, for fixed code rate 1/2, FTN signalling can increase the spectral efficiency, e.g., for $\tau = 0.8$ and $\tau = 0.6$, respectively, about 25% and 67% higher spectral efficiencies can be obtained with less than 0.2 dB BER performance loss in terms of E_b/N_0. On the other hand, if the spectral efficiency is fixed at$\rho = 0.67$, the BER performance can be improved by reducing both τ and code rate, e.g., about 0.2 dB gain can be obtained with $\tau = 0.8$ and code rate 2/5 compared with the Nyquist counterpart. However, further decreasing τ and code rate may lead to performance loss, e.g., $\tau = 0.5$ and code rate

[h]Note that the power loss due to the training sequence has been taken into account in E_b/N_0

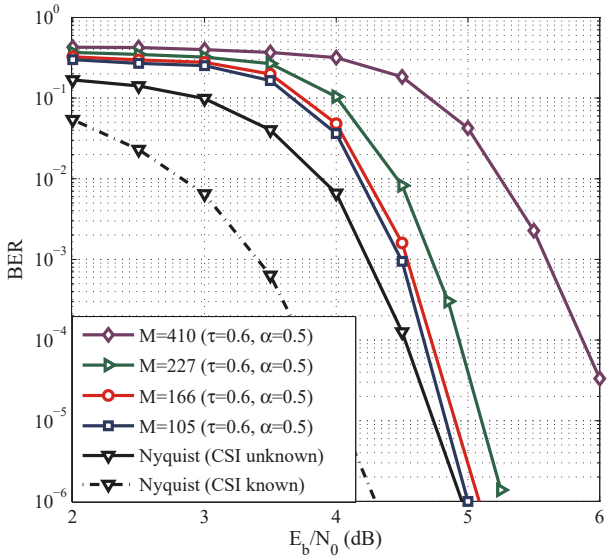

Figure 4.8: BERs of the proposed R-GAMP-JCED algorithm with different length of subblock, i.e., $M \in \{105, 166, 227, 410\}$.

1/4. This is due to the fact that the coding gain cannot be fully exploited when the ISI imposed by FTN signalling is not eliminated.

In Fig. 4.8, the BER performance of the proposed R-GAMP-JCED algorithm is investigated with different normalized fading rates. For convenience, the length of channel taps is set to 61, i.e., $L_h = 61$, and the length of the subblock $M \in \{105, 166, 227, 410\}$, i.e., $\tau M f_D T_s \in \{0.053, 0.083, 0.114, 0.205\}$. The packing ratio for FTN signalling is set to $\tau = 0.6$. It can be easily seen that the performance degrades when increasing the subblock length M. This is reasonable since the quasi-static assumption of the fading channels becomes more and more inaccurate by increasing M. Moreover, the parallel degree of the receiver can also be improved by reducing M, since more independent subblocks can be proceeded simultaneously. However, the value of M cannot be smaller than the total number of the ISI taps, i.e., M should be greater than L (i.e., 79 in this example). Also, smaller M may damage the efficient approximation of the colored noise. The MSEs of channel estimation with R-GAMP-JCED algorithm are illustrated in Fig. 4.9. Obviously, smaller M leads to more accurate channel estimation, which is consistent with the BER performance results in Fig. 4.8. For $M = 410$, an obvious gap is observed, even at large E_b/N_0. Therefore, the parameter M should be carefully designed by

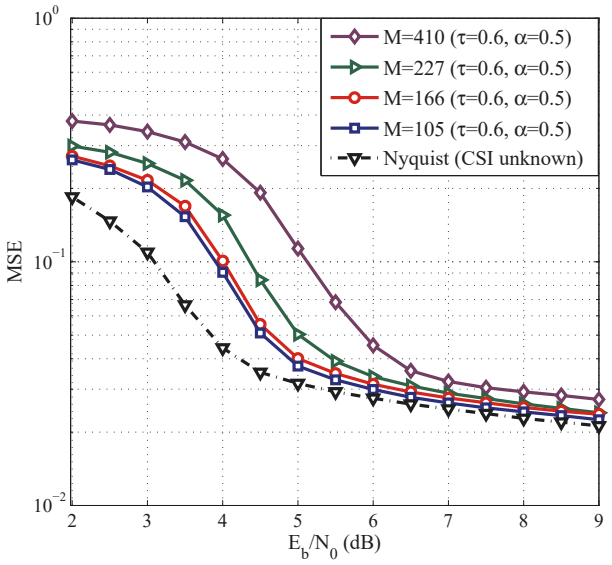

Figure 4.9: MSEs of the proposed R-GAMP-JCED algorithm with different length of subblock, i.e., $M \in \{105, 166, 227, 410\}$.

considering the normalized fading rate in practice.

4.5 Conclusions

In this chapter, frequency-domain joint channel estimation and decoding methods are developed for FTN signalling transmitting systems over frequency-selective fading channels. To deal with the colored noise inherent in FTN signalling, the corresponding autocorrelation matrix is approximated by a circulant matrix, the special eigenvalue decomposition of which facilitates an efficient fast Fourier transform operation and decoupling the noise in frequency domain. Through a specific partition of the received symbols, many independent estimates are obtained and combined to further improve the accuracy of the channel estimation and data detection. Moreover, instead of assuming the data symbols to be Gaussian random variables, a GAMP-based equalization is developed and embedded in the turbo iterations between the channel estimation and the soft-in soft-out decoder. Simulation results show that the proposed algorithm outperforms the CP-based and overlap-based frequency-domain equalization methods. With the proposed algorithms, FTN signalling reaches up to 67% higher transmission rate compared to the Nyquist

counterpart without substantially consuming more transmitter energy per bit, and the overall complexities grow logarithmically with the length of the observations.

Chapter 5

Conclusion and Future Work

This thesis presents iterative detection algorithms for high spectral efficiency communication systems operating in the high-frequency bands. In this chapter, the contributions of this thesis are summarized and future research directions are discussed.

In Chapter 2, several low-complexity joint PHN estimation and decoding algorithms are proposed for OFDM-IM operating in frequency-selective fading channels. Two models are considered for approximating the PHN process. The Wiener model exploits the time dependent statistics, while the DCT model dramatically reduces the number of unknown variables. Based on these pair of PHN models, a factor graph is constructed and the GAMP method is invoked, where the unknown non-linear transform matrix problem imposed by PHN is circumvented by using MF approximation. For comparison, the amalgamated BP-MF methods are presented for soft demodulation and for both sequential as well as for parallel message passing schedules. The performance vs. complexity of the proposed algorithms is compared to that of conventional pilot based method, the existing EKF and VI approach. The numerical results show that the proposed GAMP-DCT method has the best BER performance at a slightly increased computational complexity, as shown in Table 1. It is also observed that OFDM-IM outperforms its OFDM counterpart even for strong PHN, imperfect CSI and residual CFO. In this thesis, only the detection algorithms for single-mode-aided OFDM-IM are investigated. The multiple-mode OFDM-IM (MM-OFDM-IM) recently developed in [128] is a novel IM scheme,

where all available subcarriers within an MM-OFDM-IM block are activated for transmitting multiple distinguishable modes. Therefore, the throughput and the spectral efficiency of the system are improved compared to its single-mode-aided OFDM-IM counterpart. However, it is quite a challenge to design a low-complexity MM-OFDM-IM receiver in the presence of strong PHN, because both the modulation mode of each subcarrier and the corresponding PSK/QAM symbol drawn from that modulation constellation have to be detected while eliminating the ICI. If the proposed algorithm is extended to MM-OFDM-IM systems, the structured *a priori* pdf of MM-OFDM-IM symbol vector should be modified, since it is not sparse, which is in contrast to OFDM-IM. Moreover, given that severe ICI is imposed by the unknown PHN on MM-OFDM-IM, its BER performance relying on the proposed algorithms will have to be studied in the near future. The studies on OFDM-IM in this Chapter generally focused on point-to-point single-input single-output (SISO) systems, which can be unsuitable for some applications due to their limited spectral efficiency. More recently, MIMO transmission and OFDM-IM principles are combined to further boost the spectral and energy efficiency of the OFDM-IM scheme [27]. Therefore, the extension of detection algorithms from SISO systems to MIMO systems could be meaningful future work considering that multiple antennas are prevalent in today's wireless networks.

In Chapter 3, the amalgamated BP-MF message passing algorithm is proposed based iterative detection methods for single-carrier systems relying on DM-aided IM communicating over nonlinear dispersive satellite channels. By taking full account of the mode-selection pattern constraint, the structured *a priori* information of the Sat-DMIM signal gleaned from the SISO decoder was explicitly constructed. Based on this a factor graph was conceived by factorizing the joint *a posteriori* pdf of the information bits, the coded bits and the Sat-DMIM symbols. The BP method was applied at the mode-selection pattern constraint nodes and at the conventional PSK/QAM modulation constraint nodes, while the MF technique was invoked at the observation nodes for decoupling the Sat-DMIM symbols. To further reduce the complexity, the complicated nonlinear messages were updated by simplifying the message parameters and approximating high-order ISI terms using the first-

order Taylor series expansion method. Simulation results show that the proposed BP-MF almost achieves the FB equalizer at a lower computational complexity, as shown in Table II. Comparing with the existing linear MMSE, nonlinear MMSE, nonlinear SIC and Gaussian noise approximation based methods which are designed for conventional Satcom and then extended to the Sat-DMIM schemes, the proposed methods can efficiently improve the BER performance. The robustness of the Sat-DMIM scheme relying on the proposed methods is also demonstrated compared to its conventional Satcom and single-mode-aided Sat-IM counterparts with nonlinear distortions. In this thesis, only iteration detection methods are investigated for single-carrier systems relying on DM-aided IM in known nonlinear channels. However, the channel coefficients are always unknown in practical systems. Therefore, it is natural to study joint data detection and nonlinear channel estimation methods in the future. In addition, for the IM schemes, there are still interesting as well as challenging research problems that need to be solved in order to further improve the efficiency of IM schemes, such as the design of novel generalized/enhanced IM schemes with higher spectral and energy efficiency.

In Chapter 4, low-complexity joint channel estimation and decoding algorithms are proposed based on the VB framework for coded FTN signalling over frequency-selective fading channels. By partitioning the symbol sequence into several subblocks, a frequency-domain system model was derived, in which the interferences between subblocks due to the absence of CP were explicitly expressed and the covariance matrix of the colored noise imposed by FTN signalling was approximated by a circulant matrix. Then, the MF approach is employed to approximate the joint posterior distribution and solved the MAP symbol detection by updating the channel estimation and symbol detection iteratively. Based on the Gaussian assumption of data symbols, we derived JCED algorithm in which only the mean vector and covariance matrix have to be updated. By further exploiting the correlation between the elements in frequency domain CIRs, a refined algorithm, i.e., R-JCED was proposed to improve the channel estimation accuracy. The GAMP algorithm is further employed in which data symbols were taken as discrete variables, and proposed GAMP-JECD and R-GAMP-JCED algorithms. Simulation results

showed that, with known CIR, the proposed algorithms outperformed the existing CP-based and overlap-based MMSE-FDE methods in FTN system. When CIR is unknown, the proposed algorithms for FTN signalling performed very close to the Nyquist counterpart, while achieving up to 67% higher transmission rate with roll-off factor $\alpha = 0.5$ of raised cosine pulse without substantially consuming more transmitter energy per bit. The overall complexity of the proposed algorithms grows in order of $\mathcal{O}(N \log N)$ with the length of observations N, which is attractive for long ISI systems, e.g., FTN signalling with small packing ratio. In this thesis, only iteration detection methods are investigated for FTN systems in linear dispersive channels. Recently, the time-frequency packing transmission technique relying on both time packing of adjacent symbols and reducing the carrier spacing of adjacent channels has attracted widespread attention due to its higher spectral efficiency compared with that of time-domain FTN signalling [49]. However, the detector design for communication systems employing time-frequency packing becomes more challenging, since the receiver has to combat ISI, as well as ICI and colored noise. Therefore, how to perform data detection for systems relying on time-frequency packing in frequency-selective fading channels is worthy of further investigation.

Bibliography

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tut.*, vol. 17, no. 4, pp. 2347–2376, 2015.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[3] M. Jia, X. Gu, Q. Guo, W. Xiang, and N. Zhang, "Broadband hybrid satellite-terrestrial communication systems based on cognitive radio toward 5G," *IEEE Wireless Commun.*, vol. 23, no. 6, pp. 96–106, December 2016.

[4] Z. Lin, M. Lin, J. Wang, T. de Cola, and J. Wang, "Joint beamforming and power allocation for satellite-terrestrial integrated networks with non-orthogonal multiple access," *IEEE J. Sel. Topics Signal Process*, vol. 13, no. 3, pp. 657–670, June 2019.

[5] W. Abderrahim, O. Amin, M. Alouini, and B. Shihada, "Latency-aware offloading in integrated satellite terrestrial networks," *IEEE Open J. Commun. Society*, vol. 1, pp. 490–500, 2020.

[6] M. Giordani and M. Zorzi, "Satellite communication at millimeter waves: a key enabler of the 6G era," in *Int. Conf. Comput. Network. Commun. (ICNC)*, 2020, pp. 383–388.

[7] G. Maral, M. Bousquet, and Z. Sun, *Satellite communications systems: systems, techniques and technology.* John Wiley & Sons, 2020.

[8] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Surveys Tut.*, vol. 13, no. 4, pp. 524–540, 2011.

[9] Y. Tao, L. Liu, S. Liu, and Z. Zhang, "A survey: Several technologies of non-orthogonal transmission for 5G," *China Commun.*, vol. 12, no. 10, pp. 1–15, 2015.

[10] E. Basar, "Index modulation techniques for 5G wireless networks," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 168–175, July 2016.

[11] E. Basar, M. Wen, R. Mesleh, M. Di Renzo, Y. Xiao, and H. Haas, "Index modulation techniques for next-generation wireless networks," *IEEE Access*, vol. 5, pp. 16 693–16 746, 2017.

[12] P. Yang, M. Di Renzo, Y. Xiao, S. Li, and L. Hanzo, "Design guidelines for spatial modulation," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 6–26, Firstquarter 2015.

[13] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5536–5549, Nov 2013.

[14] M. Nakao, T. Ishihara, and S. Sugiura, "Single-carrier frequency-domain equalization with index modulation," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 298–301, Feb 2017.

[15] F.-L. Luo and C. J. Zhang, *Signal processing for 5G: algorithms and implementations*. John Wiley & Sons, 2016.

[16] J. B. Anderson, F. Rusek, and V. Öwall, "Faster-than-nyquist signaling," *Proc. IEEE*, vol. 101, no. 8, pp. 1817–1830, Aug 2013.

[17] S. Sugiura, "Frequency-domain equalization of Faster-than-Nyquist signaling," *IEEE Wireless Commun. Lett.*, vol. 2, no. 5, pp. 555–558, October 2013.

[18] T. Ishihara and S. Sugiura, "Iterative frequency-domain joint channel estimation and data detection of faster-than-Nyquist signaling," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6221–6231, 2017.

[19] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmwave) for 5G: opportunities and challenges," *Wireless networks*, vol. 21, no. 8, pp. 2657–2676, 2015.

[20] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[21] H. Huang, W. G. Wang, and J. He, "Phase noise and frequency offset compensation in high frequency MIMO-OFDM system," in *IEEE Int. Conf. Commun. (ICC)*. IEEE, 2015, pp. 1280–1285.

[22] M. Dohler, R. W. Heath, A. Lozano, C. B. Papadias, and R. A. Valenzuela, "Is the PHY layer dead?" *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 159–165, 2011.

[23] B. Li, C. Zhao, M. Sun, H. Zhang, Z. Zhou, and A. Nallanathan, "A Bayesian approach for nonlinear equalization and signal detection in millimeter-wave communications," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3794–3809, 2015.

[24] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, 2016.

[25] J. Choi, "Single-carrier index modulation for IoT uplink," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 6, pp. 1237–1248, 2019.

[26] X. Zhu, C. Jiang, L. Kuang, N. Ge, and J. Lu, "Non-orthogonal multiple access based integrated terrestrial-satellite networks," *IEEE J. Sel. Areas Commun*, vol. 35, no. 10, pp. 2253–2267, Oct 2017.

[27] M. Di Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, opportunities, and implementation," *Proc. IEEE*, vol. 102, no. 1, pp. 56–103, 2013.

[28] W.-C. Liu, T.-C. Wei, Y.-S. Huang, C.-D. Chan, and S.-J. Jou, "All-digital synchronization for SC/OFDM mode of IEEE 802.15. 3c and IEEE 802.11 ad," *IEEE Trans. Circuits and Systems I: Regular Papers*, vol. 62, no. 2, pp. 545–553, 2014.

[29] D. Tsonev, S. Sinanovic, and H. Haas, "Enhanced subcarrier index modulation (SIM) OFDM," *Proc. IEEE GLOBECOM Workshops*, pp. 728–732, Dec 2011.

[30] R. Fan, Y. J. Yu, and Y. L. Guan, "Generalization of orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5350–5359, Oct 2015.

[31] M. Wen, B. Ye, E. Basar, Q. Li, and F. Ji, "Enhanced orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4786–4801, July 2017.

[32] H. Zhang, L. Yang, and L. Hanzo, "Compressed sensing improves the performance of subcarrier index-modulation-assisted OFDM," *IEEE Access*, vol. 4, pp. 7859–7873, 2016.

[33] L. Piazzo and P. Mandarini, "Analysis of phase noise effects in OFDM modems," *IEEE Trans. Commun.*, vol. 50, no. 10, pp. 1696–1705, Oct. 2002.

[34] N. Ishikawa, S. Sugiura, and L. Hanzo, "Subcarrier-index modulation aided OFDM - will it work?" *IEEE Access*, vol. 4, pp. 2580–2593, 2016.

[35] Y. Xiao, S. Wang, L. Dan, X. Lei, P. Yang, and W. Xiang, "OFDM with interleaved subcarrier-index modulation," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1447–1450, Aug 2014.

[36] E. Basar, "OFDM with index modulation using coordinate interleaving," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 381–384, Aug 2015.

[37] O. H. Salim, A. A. Nasir, H. Mehrpouyan, W. Xiang, S. Durrani, and R. A. Kennedy, "Channel, phase noise, and frequency offset in OFDM systems: Joint estimation, data detection, and hybrid cramer-rao lower bound," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3311–3325, Sep. 2014.

[38] D. D. Lin and T. J. Lim, "The variational inference approach to joint data detection and phase noise estimation in OFDM," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1862–1874, May. 2007.

[39] R. Wang, H. Mehrpouyan, M. Tao, and Y. Hua, "Channel estimation, carrier recovery, and data detection in the presence of phase noise in OFDM relay systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1186–1205, Feb 2016.

[40] T. Ishihara and S. Sugiura, "Faster-than-Nyquist signaling with index modulation," *IEEE Wireless Commun. Lett.*, vol. 6, no. 5, pp. 630–633, Oct 2017.

[41] A. Gutierrez and W. E. Ryan, "Performance of volterra and MLSD receivers for nonlinear band-limited satellite systems," *IEEE Trans. Commun.*, vol. 48, no. 7, pp. 1171–1177, July 2000.

[42] J. P. Choi and C. Joo, "Challenges for efficient and seamless space-terrestrial heterogeneous networks," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 156–162, May 2015.

[43] L. Kuang, X. Chen, C. Jiang, H. Zhang, and S. Wu, "Radio resource management in future terrestrial-satellite communication networks," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 81–87, October 2017.

[44] M. Nakao, T. Ishihara, and S. Sugiura, "Dual-mode time-domain index modulation for Nyquist-criterion and faster-than-Nyquist single-carrier transmissions," *IEEE Access*, vol. 5, pp. 27 659–27 667, 2017.

[45] T. Mao, Z. Wang, Q. Wang, S. Chen, and L. Hanzo, "Dual-mode index modulation aided OFDM," *IEEE Access*, vol. 5, pp. 50–60, 2017.

[46] X. Li, H. Wang, N. Guan, and W. Lai, "A dual-mode index modulation scheme with gray-coded pairwise index mapping," *IEEE Commun. Lett.*, vol. 22, no. 8, pp. 1580–1583, 2018.

[47] J. E. Mazo, "Faster-than-Nyquist signaling," *Bell Syst. Technical J.*, vol. 54, no. 8, pp. 1451–1462, Oct 1975.

[48] F. Rusek and J. B. Anderson, "Constrained capacities for faster-than-Nyquist signaling," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 764–775, Feb 2009.

[49] A. Barbieri, D. Fertonani, and G. Colavolpe, "Time-frequency packing for linear modulations: spectral efficiency and practical detection schemes," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 2951–2959, October 2009.

[50] J. Fan, S. Guo, X. Zhou, Y. Ren, G. Y. Li, and X. Chen, "Faster-than-Nyquist signaling: An overview," *IEEE Access*, vol. 5, pp. 1925–1940, 2017.

[51] P. Banelli, S. Buzzi, G. Colavolpe, A. Modenini, F. Rusek, and A. Ugolini, "Modulation formats and waveforms for 5G networks: Who will be the heir of OFDM? An overview of alternative modulation schemes for improved spectral efficiency," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 80–93, 2014.

[52] J. B. Anderson, F. Rusek, and V. Owall, "Faster-than-Nyquist signaling," *Proc. IEEE*, vol. 101, no. 8, pp. 1817–1830, 2013.

[53] A. D. Liveris and C. N. Georghiades, "Exploiting faster-than-Nyquist signaling," *IEEE Trans. Commun.*, vol. 51, no. 9, pp. 1502–1511, 2003.

[54] F. Rusek and J. B. Anderson, "Multistream faster than Nyquist signaling," *IEEE Trans. Commun.*, vol. 57, no. 5, pp. 1329–1340, 2009.

[55] A. Prlja and J. B. Anderson, "Reduced-complexity receivers for strongly narrowband intersymbol interference introduced by Faster-than-Nyquist signaling," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2591–2601, September 2012.

[56] S. Sugiura and L. Hanzo, "Frequency-domain-equalization-aided iterative detection of Faster-than-Nyquist signaling," *IEEE Trans. Veh. Technol.*, vol. 64, no. 5, pp. 2122–2128, May 2015.

[57] T. Ishihara and S. Sugiura, "Frequency-domain equalization aided iterative detection of faster-than-Nyquist signaling with noise whitening," in *Proc. IEEE Int. Conf. Commun.*, May 2016, pp. 1–6.

[58] H. Fukumoto and K. Hayashi, "Overlap frequency domain equalization for Faster-than-Nyquist signaling," *arXiv preprint arXiv:1509.00562*, 2015.

[59] R. Dinis, B. Cunha, F. Ganhao, L. Bernardo, R. Oliveira, and P. Pinto, "A hybrid ARQ scheme for faster than Nyquist signaling with iterative frequency-domain detection," in *Proc. IEEE Veh. Technol. Conf. (Spring)*, May 2015, pp. 1–5.

[60] W. Yuan, N. Wu, H. Wang, and J. Kuang, "Variational inference-based frequency-domain equalization for Faster-than-Nyquist signaling in doubly selective channels," *IEEE Signal Process. Lett.*, vol. 23, no. 9, pp. 1270–1274, Sept 2016.

[61] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351–1366, May 2003.

[62] N. Wu, W. Yuan, Q. Guo, and J. Kuang, "A hybrid BP-EP-VMP approach to joint channel estimation and decoding for FTN signaling over frequency selective fading channels," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2017.

[63] N. Wu, W. Yuan, H. Wang, Q. Shi, and J. Kuang, "Frequency-domain iterative message passing receiver for Faster-than-Nyquist signaling in doubly selective channels," *IEEE Wireless Commun. Lett.*, vol. 5, no. 6, pp. 584–587, Dec 2016.

[64] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[65] J. Winn and C. M. Bishop, "Variational message passing," *Journal of Machine Learning Research*, vol. 6, no. Apr, pp. 661–694, 2005.

[66] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," *in Proc. IEEE Inf. Symp. Inf. Theory*, pp. 2168–2172, 2011.

[67] E. Riegler, G. E. Kirkelund, C. N. Manchon, M. Badiu, and B. H. Fleury, "Merging belief propagation and the mean field approximation: A free energy approach," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 588–602, Jan. 2013.

[68] X. Meng, S. Wu, L. Kuang, D. Huang, and J. Lu, "Multi-user detection for spatial modulation via structured approximate message passing," *IEEE Commun. Lett.*, vol. 20, no. 8, pp. 1527–1530, Aug 2016.

[69] L. Wei, J. Zheng, and Q. Liu, "Approximate message passing detector for index modulation with multiple active resources," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 972–976, Jan 2019.

[70] M. Tuchler, R. Koetter, and A. C. Singer, "Turbo equalization: principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, May 2002.

[71] N. Noels, J. Bhatti, H. Bruneel, and M. Moeneclaey, "Block-processing soft-input soft-output demodulator for coded PSK using DCT-based phase noise estimation," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2939–2950, Aug. 2014.

[72] J. Xu, L. Chen, I. Djurdjevic, S. Lin, and K. Abdel-Ghaffar, "Construction of regular and irregular LDPC codes: Geometry decomposition and masking," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 121–134, Jan 2007.

[73] V. Franz and J. B. Anderson, "Concatenated decoding with a reduced-search BCJR algorithm," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 186–195, Feb 1998.

[74] W. Yuan, N. Wu, A. Zhang, X. Huang, Y. Li, and L. Hanzo, "Iterative receiver design for FTN signaling aided sparse code multiple access," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2019.

[75] Q. Guo, D. Huang, S. Nordholm, J. Xi, and Y. Yu, "Iterative frequency domain equalization with generalized approximate message passing," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 559–562, Jun. 2013.

[76] Q. Shi, N. Wu, X. Ma, and H. Wang, "Frequency-domain joint channel estimation and decoding for faster-than-Nyquist signaling," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 781–795, Feb. 2018.

[77] M. D. Renzo, H. Haas, and P. M. Grant, "Spatial modulation for multiple-antenna wireless systems: a survey," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 182–191, December 2011.

[78] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228–2241, July 2008.

[79] X. Wang and B. Hu, "A low-complexity ML estimator for carrier and sampling frequency offsets in OFDM systems," *IEEE Commun. Lett.*, vol. 18, no. 3, pp. 503–506, March 2014.

[80] W. Wang, Z. Wang, C. Zhang, Q. Guo, P. Sun, and X. Wang, "A BP-MF-EP based iterative receiver for joint phase noise estimation, equalization, and decoding," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1349–1353, Oct. 2016.

[81] M. Morelli and M. Moretti, "Fine carrier and sampling frequency synchronization in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1514–1524, April 2010.

[82] S. Wu and Y. Bar-Ness, "OFDM systems in the presence of phase noise: consequences and solutions," *IEEE Trans. Commun.*, vol. 52, no. 11, pp. 1988–1996, Nov 2004.

[83] P. Aquilina and T. Ratnarajah, "Performance analysis of IA techniques in the MIMO IBC with imperfect CSI," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1259–1270, April 2015.

[84] A. Almradi and K. A. Hamdi, "Spectral efficiency of OFDM systems with random residual CFO," *IEEE Trans. Commun*, vol. 63, no. 7, pp. 2580–2590, July 2015.

[85] D. D. Lin, R. A. Pacheco, Teng Joon Lim, and D. Hatzinakos, "Joint estimation of channel response, frequency offset, and phase noise in OFDM," *IEEE Trans. Signal Processing*, vol. 54, no. 9, pp. 3542–3554, Sep. 2006.

[86] D. Ampeliotis, A. A. Rontogiannis, K. Berberidis, M. Papaleo, and G. E. Corazza, "Turbo equalization of non-linear satellite channels using soft interference cancellation," in *2008 4th Advanced Satellite Mobile Systems Conf.*, Aug 2008, pp. 289–292.

[87] G. Colavolpe and A. Piemontese, "Novel SISO detection algorithms for nonlinear satellite channels," *IEEE Wireless Commun. Lett,*, vol. 1, no. 1, pp. 22–25, February 2012.

[88] F. M. Kashif, H. Wymeersch, and M. Z. Win, "Monte carlo equalization for nonlinear dispersive satellite channels," *IEEE J. Sel. Areas Commun*, vol. 26, no. 2, pp. 245–255, February 2008.

[89] Z. Long, N. Wu, H. Wang, and Q. Guo, "Turbo equalization based on a combined VMP-BP algorithm for nonlinear satellite channels," *IEEE Access*, vol. 6, pp. 35 492–35 500, 2018.

[90] X. Wen, W. Yuan, D. Yang, N. Wu, and J. Kuang, "Low complexity message passing receiver for faster-than-nyquist signaling in nonlinear channels," *IEEE Access*, vol. 6, pp. 68 233–68 241, 2018.

[91] A. Doucet and Xiaodong Wang, "Monte Carlo methods for signal processing: a review in the statistical signal processing context," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 152–170, Nov 2005.

[92] S. Benedetto, E. Biglieri, and R. Daffara, "Modeling and performance evaluation of nonlinear satellite Links-A volterra series approach," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-15, no. 4, pp. 494–507, July 1979.

[93] C. E. Burnet and W. G. Cowley, "Performance analysis of turbo equalization for nonlinear channels," in *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.*, Sep. 2005, pp. 2026–2030.

[94] B. Benammar, N. Thomas, C. Poulliat, M. Boucheret, and M. Dervin, "On linear MMSE based turbo-equalization of nonlinear volterra channels," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2013, pp. 4703–4707.

[95] E. ETSI, "Digital video broadcasting (dvb); second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broadband satellite applications," Technical report, Tech. rep., ETSI, Tech. Rep., 2005.

[96] Lei Ding, G. T. Zhou, D. R. Morgan, Zhengxiang Ma, J. S. Kenney, Jaehyeong Kim, and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. on Commun.*, vol. 52, no. 1, pp. 159–165, 2004.

[97] Er-Wei Bai and Minyue Fu, "A blind approach to Hammerstein model identification," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1610–1619, 2002.

[98] G. Colavolpe, A. Modenini, and F. Rusek, "Channel shortening for nonlinear satellite channels," *IEEE Commun. Lett.*, vol. 16, no. 12, pp. 1929–1932, 2012.

[99] J. Zhang, C. Guo, J. Liu, X. Wu, A. P. T. Lau, C. Lu, and S. Yu, "Decision-feedback frequency-domain volterra nonlinear equalizer for IM/DD OFDM long-reach PON," *J. Lightw. Technol.*, vol. 37, no. 13, pp. 3333–3342, 2019.

[100] B. F. Beidas, "Intermodulation distortion in multicarrier satellite systems: Analysis and turbo volterra equalization," *IEEE Trans. on Commun.*, vol. 59, no. 6, pp. 1580–1590, 2011.

[101] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proc. F, Commun., Radar Signal Process.*, vol. 130, no. 1, pp. 11–16, 1983.

[102] W. Wirtinger, "Zur formalen theorie der funktionen von mehr komplexen veränderlichen," *Mathematische Annalen*, vol. 97, no. 1, pp. 357–375, 1927.

[103] B. F. Beidas, R. I. Seshadri, M. Eroz, and L. Lee, "Faster-than-Nyquist signaling and optimized signal constellation for high spectral efficiency communications in nonlinear satellite systems," in *Proc. IEEE MILCOM Conf.*, 2014, pp. 818–823.

[104] M. El-Hajjar and L. Hanzo, "EXIT charts for system design and analysis," *IEEE Commun. Surveys Tuts*, vol. 16, no. 1, pp. 127–153, 2014.

[105] X. Ma and G. B. Giannakis, "Maximum-diversity transmissions over doubly selective wireless channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1832–1840, July 2003.

[106] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, June 2003.

[107] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct 2014.

[108] K. Fang, L. Rugini, and G. Leus, "Low-complexity block turbo equalization for OFDM systems in time-varying channels," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5555–5566, Nov 2008.

[109] D. D. Lin, R. A. Pacheco, T. J. Lim, and D. Hatzinakos, "Joint estimation of channel response, frequency offset, and phase noise in OFDM," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3542–3554, Sept 2006.

[110] Q. Guo and D. Huang, "EM-based joint channel estimation and detection for frequency selective channels using Gaussian message passing," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 4030–4035, Aug 2011.

[111] H. Liu and P. Schniter, "Iterative frequency-domain channel estimation and equalization for single-carrier transmissions without cyclic-prefix," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3686–3691, October 2008.

[112] R. Koetter, A. C. Singer, and M. Tuchler, "Turbo equalization," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 67–80, Jan 2004.

[113] A. Gusmao, P. Torres, R. Dinis, and N. Esteves, "A turbo FDE technique for reduced-CP SC-based block transmission systems," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 16–20, Jan 2007.

[114] R. Dinis, P. Montezuma, L. Bernardo, R. Oliveira, M. Pereira, and P. Pinto, "Frequency-domain multipacket detection: a high throughput technique for SC-FDE systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3798–3807, July 2009.

[115] H. A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proc. IEEE*, vol. 95, no. 6, pp. 1295–1322, June 2007.

[116] P. Sen, T. Aktas, and A. O. Yilmaz, "A low-complexity graph-based LMMSE receiver designed for colored noise induced by FTN-signaling," in *IEEE Wireless Commun. Netw. Conf.*, 2014, pp. 642–647.

[117] X. Qi, N. Wu, H. Wang, and W. Yuan, "A factor graph-based iterative detection of faster-than-Nyquist signaling in the presence of phase noise and carrier frequency offset," *Digital Signal Process.*, 2016.

[118] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inform. Theory*, July 2011, pp. 2168–2172.

[119] C. Zhang, C. N. Manch, Z. Wang, and B. H. Fleury, "Message-passing receivers for single carrier systems with frequency-domain equalization," *IEEE Signal Process. Lett.*, vol. 22, no. 4, pp. 404–407, April 2015.

[120] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1462–1474, Dec 2011.

[121] H. Attias, "A variational Bayesian framework for graphical models," *Advances in Neural Information Processing Systems*, vol. 12, pp. 209–215, 2000.

[122] R. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[123] M. I. Jordan, "An introduction to probabilistic graphical models," in preparation, 2003.

[124] Q. Guo and D. Huang, "A frequency domain state-space approach to LS estimation and its application in turbo equalization," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3288–3300, July 2011.

[125] Q. Shi, N. Wu, and H. Wang, "Joint channel estimation and decoding for FTNS in frequency-selective fading channels," in *Proc. IEEE Globecom*, Dec 2016, pp. 1–6.

[126] A. Barbieri, A. Piemontese, and G. Colavolpe, "On the ARMA approximation for fading channels described by the Clarke model with applications to kalman-based receivers," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 535–540, Feb 2009.

[127] D. Chu, "Polyphase codes with good periodic correlation properties (corresp.)," *IEEE Trans. Inform. Theory*, vol. 18, no. 4, pp. 531–532, Jul 1972.

[128] M. Wen, E. Basar, Q. Li, B. Zheng, and M. Zhang, "Multiple-mode orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3892–3906, 2017.