UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

Low-complexity Iterative Receiver Design for High Spectral Efficiency Communication Systems

by

Weijie Yuan

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Sydney, Australia

2019

Statement of Originality

I, Weijie Yuan, declare that this thesis, is submitted in fulfilment of the requirements for the award of PhD in Engineering, in the School of Electrical and Data Engineering at the University of Technology Sydney. This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

The document has not been submitted for qualifications at any other academic institution.

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

This thesis is the result of a research candidature conducted with another University as part of a collaborative Doctoral degree.

Production Note:
Signature: Signature removed prior to publication.

ABSTRACT

Low-complexity Iterative Receiver Design for High Spectral Efficiency Communication Systems

by

Weijie Yuan

With the rapid development of the modern society, people have an increasing demand of higher data rate. Due to the limited available bandwidth, how to improve the spectral efficiency becomes a key issue in the next generation wireless systems. Recent researches show that, compared to the conventional orthogonal communication systems, the non-orthogonal system can transmit more information with the same resources by introducing non-orthogonality. The non-orthogonal communication systems can be achieved by using faster-than-Nyqusit (FTN) signaling to transmit more data symbols in the same time period. On the other hand, by designing appropriate codebook, the sparse code multiple access (SCMA) system can support more users while preserving the same resource elements. Utilisation of these new technologies leads to challenge in receiver design, which becomes severer in complex channel environments. This thesis studies the receiver design for high spectral efficiency communication systems. The main contributions are as follows:

1. A hybrid message passing algorithm is proposed for faster-than-Nyquist, which solves the problem of joint data detection and channel estimation when the channel coefficients are unknown. To fully exploit the known ISI imposed by FTN signaling, the interference induced by FTN signaling and channel fading are intentionally separated.

2. Gaussian message passing and variational inference based estimation algorithms are proposed for faster-than-Nyquist signaling detection in doubly selective channels. Iterative receivers using mean field and

Bethe approximations based on variational inference framework are proposed. Moreover, a novel Gaussian message passing based FTN signaling detection algorithm is proposed.

3. An energy minimisation based SCMA decoding algorithm is proposed and convergence analysis of the proposed algorithm is derived. Following optimisation theory and variational free energy framework, the posterior distribution of data symbol is derived in closed form. Then, the convergence property of the proposed algorithm is analysed.

4. A stretched factor graph is designed for MIMO-SCMA system in order to reduce the receiver complexity. Then, a convergence guaranteed message passing algorithm is proposed by convexifying the Bethe free energy. Finally, cooperative communication methods based on belief consensus and alternative direction method of multipliers are proposed.

5. A low complexity detection algorithm is proposed for faster-than-Nyquist SCMA system, which enables joint channel estimation, decoding and user activity detection in grant-free systems. The combination of FTN signaling with SCMA to further enhance the spectral efficiency is first considered. Then, a merging belief propagation and expectation propagation algorithm is proposed to estimate channel state and perform SCMA decoding.

Acknowledgements

The accomplishment of my PhD thesis is owed to the contributions and supports of many people. First of all, I would like to express my deepest thanks to my supervisor Prof. Xiaojing Huang for his guidance and support throughout my PhD study years at UTS. I would like to thank Prof. Huang for his encouragement and mentorship during discussions on doing research and paper writing. I also want to thank my co-supervisor A/Prof. Andrew Zhang for his wonderful support and valuable advice for my research. From them, I have learned to be patient and dedicated to the work, which will undoubtedly influence my life.

I would like to thank Dr. Nan Wu from Beijing Institute of Technology for his guidance in my early graduate years, from whom I have learned how to think rigorously and how to solve research problems. I am also grateful to Prof. Yonghui Li from University of Sydney and A/Prof. Qinghua Guo from the University of Wollongong for kindly inviting me to visit their groups and conduct collaborative research works. Collaborating with them has broadened my horizon and improved my research.

I am greatly thankful to Prof. Franz Hlawatch from Vienna University of Technology and Dr. Bernard Etzlinger from Johannes Kepler University Linz for their help during my two-month visit in Austria. And I also thank Prof. Lajos Hanzo from the University of Southampton, Prof. Bernard Henri Fleury from Aalborg University and Prof. Jingming Kuang form Beijing Institute of Technology for research discussions and several fruitful works.

I would like to sincerely thank all my colleagues for helpful discussions during my graduate years, especially Hao, Yijiang, Yifeng, Bin, Qiaolin, Shaoang, Shizhe, Rui, Chongwen, Tuyen, Thomas and Renè. I learned a lot from them and I really enjoyed working with them. Moreover, I want to thank the Faculty and School for

providing a warm and cozy family for me.

Finally, I would like to express my gratefulness to my parents Chenggang and Chongying for their unconditional love and confidence.

List of Publications

Published Journal Papers

J-1. **W. Yuan**, N. Wu, B. Etzlinger, H. Wang, and J. Kuang, "Cooperative Joint Localization and Clock Synchronization Based on Gaussian Message Passing in Asynchronous Wireless Networks," *IEEE Transactions on Vehicular Technology,* vol. 65, no. 9, pp. 7258-7273, Sep. 2016.

J-2. **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 Processing Letters,* Vol. 23, No. 9, pp. 1270-1274, Sep. 2016.

J-3. **W. Yuan**, N. Wu, Q. Guo, Y. Li, C. Xing, and J. Kuang, "Iterative Receivers for Downlink MIMO-SCMA: Convergence-guaranteed Message Passing and Distributed Cooperative Detection," *IEEE Transactions on Wireless Communications,* vol. 17, no. 5, pp. 3444-3458, May. 2018.

J-4. **W. Yuan**, N. Wu, B. Etzlinger, H. Wang, and J. Kuang, "Expectation Maximization-Based Passive Localization with Asynchronous Receivers: Centralized and Distributed Implementations," *IEEE Transactions on Communications,* vol. 67, no. 1, pp. 668-681, Jan. 2019.

J-5. **W. Yuan**, N. Wu, C. Yan, Y. Li, X. Huang, and L. Hanzo, "A Low-Complexity Energy Minimization Based SCMA Detector and Its Convergence Analysis," *IEEE Transactions on Vehicular Technology,* vol. 67, no. 12, pp. 12398-12403, Dec. 2018.

J-6. N. Wu, **W. Yuan**, H. Wang, and J. Kuang, "TOA-based Passive Localization of Multiple Targets with Inaccurate Receivers Based on Belief Propagation on Factor Graph," *Digital Signal Processing*, no. 49, pp. 14-23, Apr. 2016.

J-7. N. Wu, **W. Yuan**, Q. Shi, H. Wang, and J. Kuang, "Frequency-Domain Iterative Message Passing Receiver for Faster-than-Nyquist Signaling in Doubly Selective Channels," *IEEE Wireless Communications Letters,* vol. 5, no. 6, pp. 584-587, Dec. 2016.

J-8. 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. 5, pp. 6849-6858, May. 2017.

J-9. 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. 68233-68241, Oct. 2018.

J-10. Q. Shi, N. Wu, H. Wang, and **W. Yuan**, "Joint Channel Estimation and Decoding in the Presence of Phase Noise over Time-selective Flat-fading Channels," *IET Communications,* vol.10, no. 5, pp. 577-585, Nov. 2015.

J-11. 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 Processing,* no. 63, pp. 25-34, Apr. 2017.

Submitted Journal Papers

J-12. **W. Yuan**, N. Wu, Q. Guo, X. Huang, Y. Li, and L. Hanzo, "TOA Based Passive Localization Constructed over Factor Graphs: A Unified Framework," Submitted to *IEEE Transactions on Communications.*

J-13. **W. Yuan**, N. Wu, B. F. Fleury, X. Huang, and Y. Li, "A Merging Message Passing Algorithm for Cooperative Joint Localization and Synchronization," To be submitted to *IEEE Internet of Things Journal.*

J-14. **W. Yuan**, N. Wu, A. Zhang, X. Huang, Y. Li, and L. Hanzo, "Iterative Receiver Design for Faster than NqyuistSignaling-Spare Code Multiple Access System," Submitted to IEEE Transactions on Wirless Communications.

Contents

List of Figures

Abbreviation

1G	The First Generation
2G	The Second Generation
3G	The Third Generation
4G	The Fourth Generation
5G	The Fifth Generation
ADMM	Alternating Direction Method of Multipliers
AMP	Approximate Message Passing
AR	Auto Regressive
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BP	Belief Propagation
BS	Base Station
CDMA	Code Division Multiple Access
CS	Compressive Sensing
CSI	Channel State Information
dB	Decibel
DFT	Discrete Fourier Transform
DSC	Doubly Selective Channel
EM	Expectation Maximisation
EP	Expectation Propagation
EXIT	Extrinsic Information Transfer
FBMC	Filter Bank Multi Carrier
FDMA	Frequency Shaped Sliding Mode Control

FDE	Frequency Domain Equalisation
FTN	Faster Than Nyquist
GMP	Gaussian Message Passing
GSM	Global System for Mobile Communication
IDMA	Interleave Division Multiple Access
IDFT	Inverse Discrete Fourier Transform
ISI	Inter Symbol Interference
KLD	Kullback-Leibler Divergence
LDPC	Low Density Check Code
LDS	Low Density Signature
LLR	Log Likelihood Ratio
LMMSE	Linear Minimum Mean Squared Error
LS	Least Square
MAP	Maximum *A Posteriori*
MIMO	Multiple Input Multiple Output
MF	Mean Field
MMSE	Minimum Mean Squared Error
MPA	Message Passing Algorithm
MPSK	M-ary Phase Shift Keying
MUSA	Multi-User Shared Access
NOMA	Non-Orthogonal Multiple Access
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OMA	Orthogonal Multiple Access
PDMA	Pattern Division Multiple Access
PMF	Probability Mass Function
p-NOMA	power domain NOMA

QAM	Quadrature Amplitude Modulation
SCM	Single-Carrier Modulation
SCMA	Sparse Code Multiple Access
SIC	Successive Interference Cancellation
SPA	Sum Product Algorithm
TDMA	Time Division Multiple Access
VMP	Variational Message Passing

Nomenclature and Notation

Capital boldface letters denote matrices.

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

Lower-case boldface alphabets denote column vectors.

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

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

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

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

I_n is the identity matrix of dimension $n \times n$.

$\propto$ represents equality up to a constant normalization factor.

$\mathbb{E}$ is the expectation operator.

C denotes a constant.

$\mathbb{B}^K$ denotes a K-dimensional binary space.

$\mathbb{C}^K$ denotes a K-dimensional complex space.

$\mathbb{R}^{K \times K}$ denotes a $K \times K$-dimensional complex space.

$*$ denotes the convolution operator.

Chapter 1

Introduction

This chapter first presents the evolution of mobile communications and the key problems in the future wireless communication systems. Then the technologies of faster-than-Nyquist (FTN) signaling and sparse code multiple access (SCMA) are reviewed. Finally, the objectives of this thesis are summarised and the outline is given.

1.1 Background

Since 1970s, mobile communication technology has evolved rapidly, which profoundly changes the work and lifestyle of human beings and promotes economic development. In 1978, the Bell Labs designed a cellular mobile communication network using analog technology and frequency division multiple access (FDMA) [1]. This communication network, also known as the first generation mobile communication system (1G), effectively solved the capacity requirements at that time. After that, in order to overcome the problem of low call quality in 1G era, the second generation mobile communication system (2G) based on digital voice communication was proposed [2]. The new technology employed in 2G is time division multiple access (TDMA) [3]. With the increasing demand for high-speed data transmission, the third generation mobile communication system (3G) using code division multiple access (CDMA) technology [4] came into being. Relying on more bandwidth and high data rate, 3G can provide users with more wireless applications. Nevertheless, it still has some limitations in mobile multimedia services [5]. The fourth generation mobile communication system (4G) using orthogonal frequency division multiple

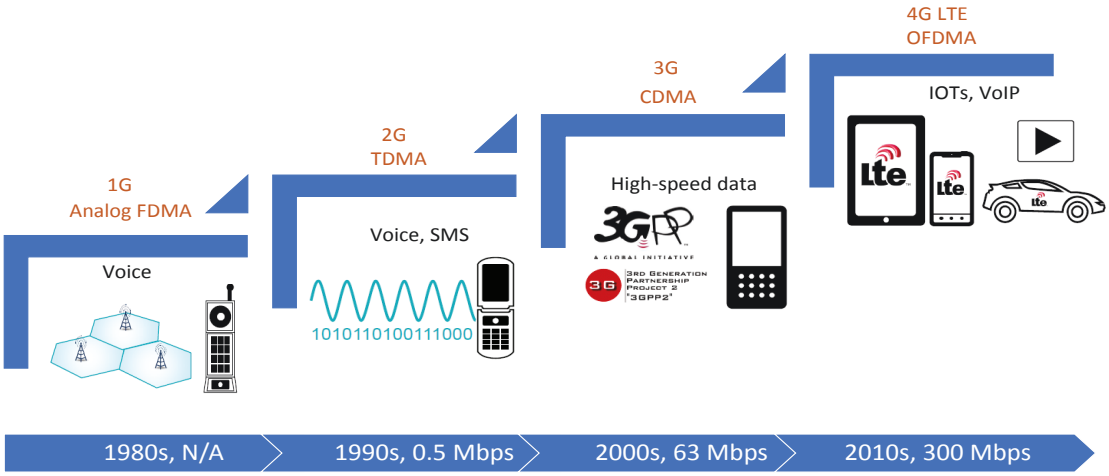


Figure 1.1 : The evolution of mobile communications.

access (OFDMA) [6] and multiple input multiple output (MIMO) [7] can address the signal coverage problem in 3G. At present, various new mobile Internet-based services have emerged that greatly improved the informationisation of the society. In Fig. 1.1 (From ref. [6]), the evolution of mobile communications from 1G to 4G is illustrated.

The history of mobile communication over the last four decades suggests that the research on new mobile communication system will never stop. On one hand, the rapid development of integrated circuit and chip design has greatly enhanced the performance of mobile devices, which enables the implementation of many complex technologies such as low density partial check (LDPC) code [8]. One the other hand, the demand for higher quality wireless services continues growing [9]. In the last ten years, the explosive growth of smartphones led to an exponential increase of the amount of mobile data [10]. At the same time, the applications including Internet of things, smart home and virtual reality requires more accessed device, higher coverage, lower transmission delay and smaller communication overhead [11–14]. According to the 5G white paper, by 2020, the next generation mobile networks should satisfy the following performance indicators: 1)transmission rate increased by

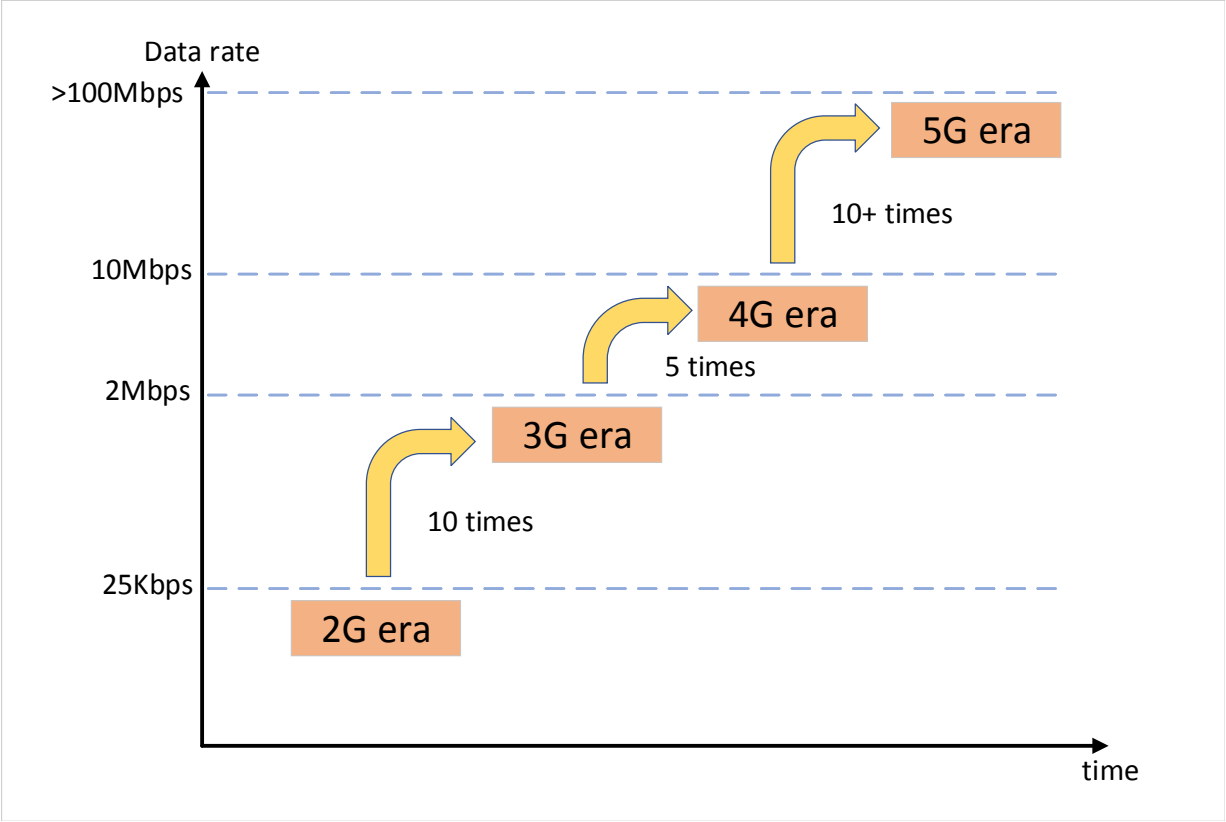


Figure 1.2 : The variation of data rate from 1G to 5G.

10-100 times; 2)capacity enhanced by 1000 times of current network capacity; 3)end-to-end delay reduced to 1/10 of the current one [15]. Fig. 1.2 shows the variation in transmission rate as the mobile communication systems evolve. Undoubtedly, the birth of 5G will further change our lives.

In order to achieve the above objectives, academia and industry have conducted research from three different points of views. The first one is to use new technology to further improve the spectral efficiency with limited bandwidth [16–18]. The second one is to find new available spectrum resources based on mm-wave and visible light communications [19–22]. The last one is to increase the system throughput by using communications between users [23–25]. Higher spectral efficiency can be achieved by designing new modulation formats and new multiple access systems. There exist several new modulation methods, e.g., FBMC, FTN signaling and SCM

that can improve the spectral efficiency [26–30]. By introducing intentional inter symbol interference (ISI), FTN signaling can transmit more data using the remaining bandwidth in 4G system [31]. For the multiple access technologies, 5G mainly considers the use of non-orthogonal multiple access (NOMA) to support more users using the same resources [32, 33]. Existing NOMA technologies include power domain NOMA (p-NOMA) [34], interleave division multiple access (IDMA) [35, 36], multiuser shared access (MUSA) [37], pattern division multiple access (PDMA) [38] and sparse code multiple access (SCMA) [39,40]. Amongst them, SCMA has received extensive attention because of its extra shaping gain. Moreover, the environment will introduce various interferences, such as multi path effect in indoor environment and time selective fading on high-speed trains. How to tackle these challenges while realising high spectral efficiency is one of the key goals in 5G. In this thesis, the low complexity reception methods for FTN signaling and SCMA systems are studied.

1.2 Faster-than-Nyquist Signaling

According to Nqyuist theorem, the maximum symbol rate should be twice the channel bandwidth. When the symbol rate equals the Nyquist rate, a sample of the matched filter output only depends on one transmitted symbol impulse. That is to say the received sample contains all the necessary information for symbol decision and symbol-by-symbol detection is used to obtain all transmitted information. Although Nyquist criterion ensures ISI-free transmission, it wastes certain spectrum resources to keep the orthogonality of waveforms [29]. In order to take advantage of these spare resources, more data symbols can be transmitted in the same time period, which is the basic principle of FTN signaling. In 1970, the concept of FTN was first introduced [41]. Then Mazo detailedly analysed FTN signaling and proved that FTN signaling can transmit 25% more information bits than Nyquist signaling in additive white Gaussian noise (AWGN) channel, while preserving the bit error

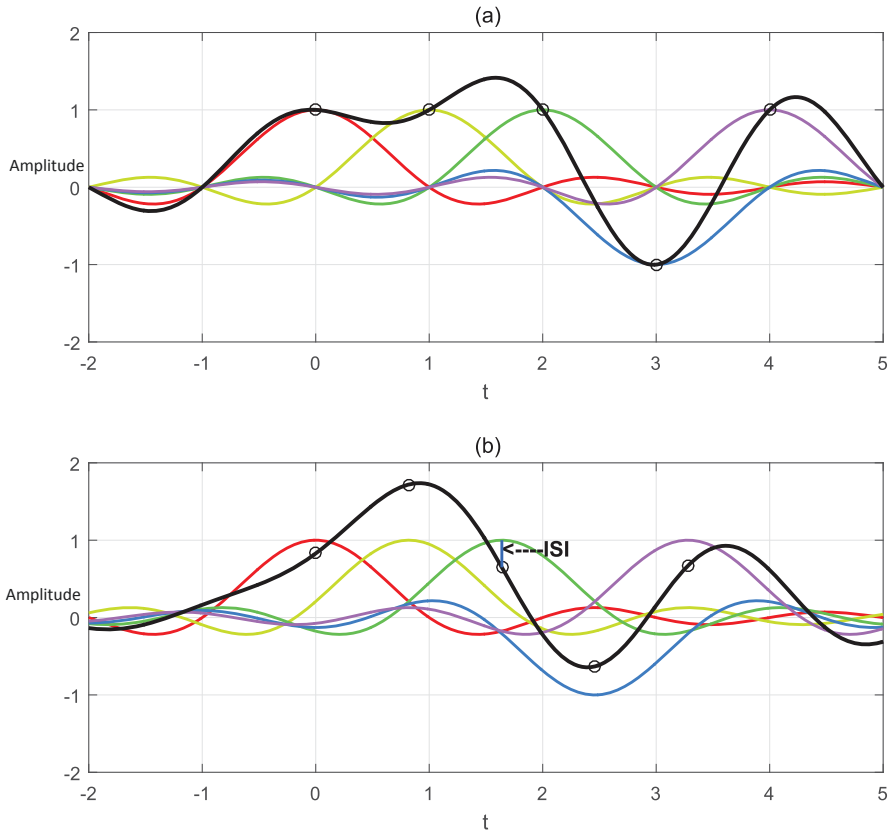


Figure 1.3 : Comparison of (a) Nyquist signaling and (b) FTN signaling.

rate (BER) performance [28].

Assuming that c_n is the transmitted symbol at time n and $q(t)$ is a shaping pulse, the Nyquist signaling can be expressed as

$$s(t) = \sum_n c_n q(t - nT_0). \tag{1.1}$$

Considering the same shaping pulse $q(t)$, the FTN signaling model reads,

$$s(t) = \sum_n c_n q(t - n\tau T_0). \tag{1.2}$$

It can be seen when $\tau = 1$, (1.2) and (1.1) are identical. In FTN signaling, setting $\tau < 1$ can transmit more information with the same bandwidth and time period, therefore τ is named as packing factor. Fig. 1.3 illustrates the Nyquist

signaling waveform and FTN signaling waveform with $\tau = 0.8$. It can be observed that in the same time period, FTN signaling can transmit more sinc pulses. It can be also observed that when sampling with symbol interval τT_0, the samples contain information of other symbols, which means that the interference between transmitted symbols should be cancelled when detecting FTN signals.

For the signal defined by (1.2), the error probability when performing detection is determined by the minimum distance of signals [42]. Considering two signals $s_i(t)$ and $s_j(t)$ given by (1.2), the minimum Euclidean distance d^2_{min} is

$$d^2_{min} = \left(\frac{1}{E_b}\right) \int_{-\infty}^{\infty} |s_i(t) - s_j(t)|^2 \mathrm{d}t, \quad i \neq j, \tag{1.3}$$

where E_b is the signal energy per bit. Assuming the signal transmits through an AWGN channel, the symbol error rate of optimal detection is given by function $Q(\sqrt{\frac{d^2_{min} E_b}{N_0}})$, where the Q-function is the tail distribution function of the standard normal distribution, formulated as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u. \tag{1.4}$$

For binary symbol Nyquist signaling, the minimum Euclidean distance of signal is always 2. The corresponding symbol error rate $Q(\sqrt{\frac{2_{min} E_b}{N_0}})$ is referred to as the matched filter bound [43]. For FTN signaling using sinc pulse, the minimum distance will not drop immediately for $\tau < 1$. In fact for binary symbol, although ISI is induced by FTN signaling, when τ is larger than 0.802, $d^2_{min} = 2$ still holds. Hence, $\tau = 0.802$ is known as the Mazo limit [28]. When the packing factor is above the Mazo limit, it is capable of increasing symbol rate while not affecting the BER performance. Subsequently, research on root raised cosine shaping pulse showed that with roll-off factor $\alpha = 0.3$, Mazo limit can be reduced to 0.703, which shows the data rate can be as 1.43 times high as Nyquist signaling [29]. According to [44], Mazo limit applies to non-binary transmissions, nonlinear pulses and even nonlinear modulations.

Based on these works, it is seen FTN signaling is very attractive in future mobile communications. However, due to the non-orthogonality of waveform, FTN signaling introduces severe ISI. It can be seen from the signal model (1.2) that each symbol is interfered by several neighboring transmitted symbols. Consequently, designing receivers for FTN signaling in order to recover original transmitted data is very important. Next, a basic discrete time FTN signaling model is considered. The transmitted signal $s(t)$ passes through an AWGN channel and is received at the receiver side. After matched filtering, the received signal $r(t)$ can be expressed as

$$r(t) = \sum_{n-\infty}^{\infty} c_n g(t - n\tau T_0) + \gamma(t), \tag{1.5}$$

where $g(t)$ is the convolution of shaping filter and matched filter and $\gamma(t)$ is the random noise process. After sampling using symbol period τT_0, the received samples $\{r_n\}$ are obtained. The objective of FTN signaling receiver is to recover sequence $\{c_n\}$ from the received signal samples $\{r_n\}$. Since the shaping pulse is non-orthogonal with repsect to symbol period τT_0, the noise samples at the output of the matched filter have correlations [29]. Usually a post filter is employed to whiten the noise [45] and then signal detection is performed. Assuming that the discrete time sequence obtained from $g(t)$ is denoted by $\{g_n\}$, FTN signaling can be regarded as the trellis code of $\{c_n\}$ by $\{g_n\}$. Detection of $\{c_n\}$ is equivalent to decoding of the noisy trellis coded sequence. Viterbi algorithm [46] can be employed to find the most likely sequence $\{c_n\}$ by comparing the received signal and the shortest path. Viterbi algorithm provides the maximum likelihood estimate of transmitted symbol. When the symbols are with equal probability, the maximum likelihood based receiver is optimal. However, when the probabilities of symbols are unequal, the maximum *a posteriori* estimator should be employed, given by

$$\hat{c}_n = \arg\max_{c_n} \frac{p(c_n|\{r_n\})}{p(\{r_n\})}. \tag{1.6}$$

Considering symbol by symbol MAP estimator, Bahl et al. proposed BCJR algorith-

m for symbol detection [47], which is also based on trellis diagram. By computing the likelihood probability of $\{c_n\}$, the transmitted sequence is detected iteratively. If a large value of τ is chosen, the FTN induced ISI is slight, then using Viterbi algorithm or BCJR algorithm is effective in symbol detection. However, when ISI becomes severer, the larger number of trellis states makes the receiver complexity prohibitively high.

To reduce the complexity of trellis based receivers, three approaches were proposed, i.e., channel shortening [48], reduced search [49] and reduced trellis [50]. Channel shortening aims for equivalising the FTN induced ISI to an ISI channel with fewer taps in order to reduce the complexity. Reduced search method only considers searching the optimal path in part of the trellis diagram. For example, the M-algorithm only considers M paths in the diagram [51]. Reduced trellis method can reduce the total number of trellis states by reducing the number of states of a symbol [50]. For FTN signaling, an M-algorithm based BCJR receiver was proposed in [52] to eliminate the ISI introduced by FTN signaling. Although only M-paths are considered, the complexity of the receiver in [52] still increases exponentially.

Based on frequency domain equalisation (FDE) method for conventional frequency selective fading channels, an FDE based minimum mean squared error (MMSE) algorithm was proposed in [53]. In FDE, time domain signal is transformed to frequency domain through faster Fourier transform (FFT), then ISI becomes frequency channel coefficients and FTN symbols are detected with linear complexity. Nevertheless, FDE has to insert cyclic prefix symbols in the transmitted sequence to convert linear convolution into circular convolution, which decreases the spectral efficiency. In time domain equalisation, the authors in [54] employed AR process to model the colored noise and proposed a linear MMSE FTN signaling receiver.

Current research on receiver design for FTN signaling mainly focuses on AWGN

channel. In practical environments, the channel suffers from multipath effect and Doppler spread, which lead to frequency and time selective fadings. The interference induced by fading channels makes the FTN signaling receiver more complex. Moreover, when the channel information is unknown, how to achieve good detection performance of FTN symbols is still under investigation. To this end, this thesis will study the joint FTN signaling detection and channel estimation over frequency selective channels and FTN data detection in doubly selective channels.

1.3 Sparse Code Multiple Access

In conventional orthogonal multiple access (OMA) technology, each user will be assigned to certain orthogonal radio resource elements, such as orthogonal frequency division multiplexing (OFDM) sub-carriers and MIMO antennas. So the number of users cannot exceed the total number of orthogonal resource elements. At the receiver side, the signals corresponding to different users can be obtained through simple single user detection. When the channel condition of a user is poor, OMA has to set high priority to transmit this user's data in order to satisfy fairness. This results in both the waste of spectrum and the decline of system throughput. Theoretically, OMA cannot achieve the sum-rate capacity of multiuser systems. To this end, non-orthogonal resource allocation is proposed in multiple access system, which is the NOMA technology. By introducing interference between users, one resource element in NOMA can support more users to transmit information, and the increment of the information rate makes the improvement of spectral efficiency. The multiuser capacity in AWGN channel is analyzed in [55], which demonstrated that NOMA can reach the capacity bound. In multipath fading channels, NOMA achieves the optimal capacity when the channel information is only known at the receiver side, while OMA is strictly suboptimal. Moreover, since more users are supported by NOMA, the requirement of massive connectivity in 5G can be satisfied [32].

The NOMA technologies can be categorised into power domain methods and code domain methods [32]. The power domain NOMA (p-NOMA) maximises the system gain by assigning different power levels to users according to their channel conditions. This enables us to distinguish different users at the receiver side, then use successive interference cancellation (SIC) to detect the signals. In code domain NOMA technologies, the users will be assigned different codewords to achieve multiplexing. Compared to p-NOMA, code domain NOMA technologies can achieve spreading gain and higher sum rate [56].

SCMA technology can be regarded as the extension of the low density signature (LDS) method [39]. Different from LDS technology, SCMA maps bit streams of different users to SCMA codewords directly, which makes joint codebook and constellation optimisation possible. Fig. 1.4 illustrates the SCMA encoding process of a 6-user, 4-resource SCMA system. Each user maps its bits to an SCMA codeword chosen from a predefined SCMA codebook and then multiplexed over 4 resource elements. Multiuser detection is performed at the receiver side to determine the bit sequences of users. In SCMA system, a binary vector $\mathbf{f}_k$ is employed to indicate the resource elements occupied by user k. The jth element in $\mathbf{f}_k$ is defined as

$$f_{k,j} = \begin{cases} 0 & x_{k,j} = 0 \\ 1 & x_{k,j} \neq 0. \end{cases} \tag{1.7}$$

By stacking $\mathbf{f}_k$, we have the indicator matrix $\mathbf{F} = [\mathbf{f}_1, ..., \mathbf{f}_K]$. In $\mathbf{F}$, the non-zero entries in the jth row denote the conflicting users over the jth antenna while the non-zero entries in the kth column indicate the resources occupied by user k. The indicator matrix corresponding to the SCMA system shown in Fig. 1.4 is given as

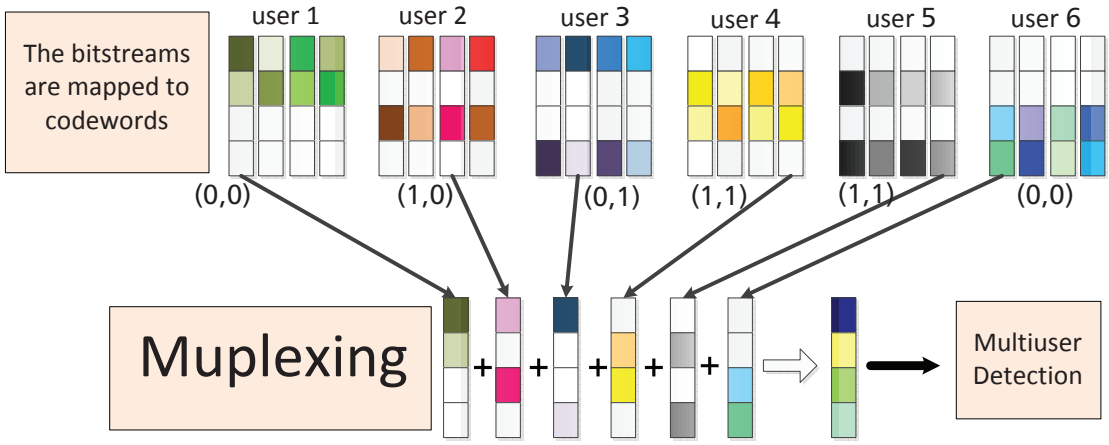


Figure 1.4 : SCMA encoding process.

follows,

$$\mathbf{F} = \begin{bmatrix} \mathbf{1} & 1 \\ \mathbf{1} & 0 \\ \mathbf{0} & 1 \\ \mathbf{0} & 0 \end{bmatrix}. \tag{1.8}$$

It is observed from $\mathbf{F}$ that when designing the codebook, the 0 valued positions of different users' codewords are different, which efficiently avoid packet collisions. Besides, only partial resource elements are assigned to one user, which ensures one resource element will support a few users and the sparsity property makes the complexity of SCMA system still controllable [57].

In the overloaded SCMA system, the received signal contains the interference introduced by other users and this should be taken into account when designing receivers. For the indicator matrix $\mathbf{F}$ in (1.8), a factor graph can be used to show the relationship between the users and resource elements [58], as shown in Fig. 1.5. The factor graph contains factor vertices E and variable vertices U, where a variable vertex represents one user's transmitted symbols and a factor vertex represents the function relationship of the source elements and transmitted symbols. If and only if a

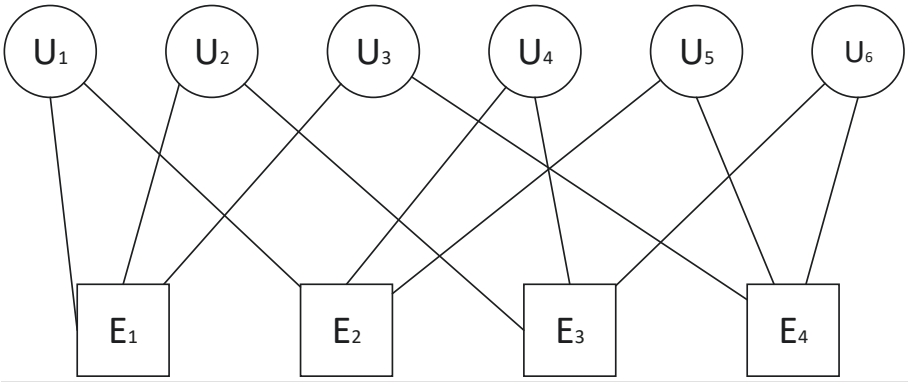


Figure 1.5 : The factor graph representation of SCMA system.

user occupies a resource element, the variable vertex and factor vertex are connected by an edge. Thanks to the sparsity of SCMA codewords, belief propagation (BP) can be used on factor graph to derive the MAP estimate of transmitted symbol. On the factor graph, there are two kinds of messages, namely, the message from variable vertex x to factor vertex f and the message from factor vertex f to variable vertex x, denoted by $\mu_{f\to x}(x)$ and $\mu_{x\to f}(x)$, respectively. According to the BP update rules [58], $\mu_{f\to x}(x)$ and $\mu_{x\to f}(x)$ are given by

$$\mu_{f\to x}(x) \propto \int f(\mathbf{x}) \prod_{x^{'}\in\mathcal{N}(f)\setminus\{x\}} \mu_{x^{'}\to f}(x^{'})\mathrm{d}x^{'}, \tag{1.9}$$

$$\mu_{x\to f}(x) \propto \prod_{f^{'}\in\setminus\{f\}} \mu_{f^{'}\to x}(x), \tag{1.10}$$

where $\mathcal{N}(x)$ and $\mathcal{N}(f)$ denote the sets of all factors connected to x and all variables in the function f, respectively. And the belief (approximate marginal) of variable x can be expressed as

$$b(x) \propto \prod_{f\in\mathcal{N}(x)} \mu_{f\to x}(x). \tag{1.11}$$

By passing messages on factor graph, the marginal distribution of SCMA symbol is determined iteratively. From (1.9), it can be observed that the message calculation requires the integration over all other interfered symbols. Therefore the complex-

ity of the conventional BP algorithm increases exponentially with the number of interfered symbols.

To reduce the complexity of conventional BP receiver, several modified BP message passing receivers are proposed for SCMA system. In [59], the authors proposed a shuffled message passing algorithm to accelerate the convergence. [60] presented a fixed low complexity detector for uplink SCMA system based partial marginalisation. In [61], a Monte Carlo Markov Chain (MCMC) based SCMA decoder was proposed which features low complexity when the codebook size is large. In [62], the authors compute the messages in log-domain and then the multiplication operations become simple summations.

In a word, many researches have proposed low complexity receiver designs for SCMA system. However, there are still a few points that should be addressed. The first problem is the symbol detection in complex environment. Existing works consider very simple channel model and assume that the channel state information is perfectly known, which ignores complex channel conditions in practice. The second problem is the convergence of BP algorithm. Although, [63] showed that a loopy BP is still efficient in a factor graph with cycles, the convergence problem may result in performance loss. The third one is the combination of NOMA and non-orthogonal waveform. SCMA technology and FTN signaling can improve the spectral efficiency from different degrees. Naturally, a combination of both technologies is expected to achieve ever higher spectral efficiency. To tackle above challenges, this thesis will design low complexity receivers for SCMA systems in complex environments. Moreover, considering a system combining FTN signaling and SCMA technology, low complexity and communication overhead detection methods will be developed.

1.4 Thesis Objectives and Organisation

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

i. The channel state information is assumed to be known in existing FTN signaling receivers. When the multipath channel is unknown, how to tackle the combined ISI induced by fading channels and FTN signaling and then estimate the channel coefficients is very important. It is highly demanded to design new reception algorithms that jointly estimate channel taps and detect FTN data symbols.

ii. The FDE-based receivers can tackle the ISI introduced by FTN signaling. However, in high mobility environments, due to the time-variant channel, existing FDE-based algorithms are not available. Therefore, it is necessary to design novel low complexity FDE-based algorithm to eliminate the interferences caused by time selectivity and improve the robustness of FTN signaling in high mobility environments.

iii. Conventional BP receivers for SCMA system may suffers from convergence problem since BP does not guarantee convergence on loopy graphs. Therefore, the convergence analysis of SCMA receiver is necessary. Moreover, in complex environment, severer interference results in a factor graph with more short cycles. To tackle the problem that BP fails to converge, it is necessary to design a convergence-guaranteed message passing algorithm for this scenario.

iv. In multiuser system, cooperative detection can be performed by users to achieve diversity gain. The basic idea to share the measurements among all users is not realistic due to power consuming. Therefore, to develop new cooperative detection schemes with low cost is necessary.

v. A combination of SCMA and FTN technologies is expected to improve the spectral efficiency of communication systems at the cost of high complexity receiver. Moreover, in uplink system, communication overhead at the base station can be reduced by detecting the users' activities. Therefore, a novel

low complexity receiver aiming for detecting active users, decoding and channel estimation should be designed.

This thesis is organised as follows:

- *Chapter 1*: Chapter 1 presents the research background of this thesis and introduces the FTN signaling and SCMA technology. Then the objectives and organisation of this thesis are given.

- *Chapter 2*: Instead of mixing the ISI imposed by FTN signaling with the frequency selective channel together, they are intentionally separated from each other, which enables us to fully exploit the known structure of the FTN-induced ISI. Considering that the packing symbol period will lead to correlation of channel taps, discrete Fourier transform (DFT) interpolation is used to obtain channel taps in FTN scenario. The colored noise due to the sampling rate in FTN system is approximately modeled by autoregressive (AR) process to avoid the whitening filtering. Building on this, a Forney style factor graph is constructed and GMP is employed to update messages on the graph. Considering that the inner product between FTN symbols and channel coefficients is infeasible by using BP, it is proposed to perform VMP on an equivalent soft node instead. Simulation results show that the proposed algorithm can estimate the channel coefficients and detect data symbols efficiently.

- *Chapter 3*: This chapter studies the iterative receiver design for FTN signaling over doubly selective channels. The received FTN signaling samples over doubly selective channel are first modeled. Then the FDE-MMSE algorithm is extended to doubly selective channel case but shows its inadequacy in this condition. Next, starting from the variational inference and Gaussian message passing (GMP) frameworks, two novel receiver structures are proposed. In the

variational inference-based receiver design, since mean filed approximation will lead to the underestimation of variance, Bethe approximation is used to design the receiver. For the GMP based receiver, by implementing message passing on Forney-style factor graph, the extrinsic information of data symbols are efficiently calculated. Moreover, considering that the channel state information is not perfectly known, the factor graph is extended and a robust receiver is developed. Simulation results show both receivers for FTN signaling can achieve similar performance of Nyquist signaling.

- *Chapter 4*: In Chapter 4, a low-complexity receiver based on Bayesian inference is proposed. An uplink SCMA model is first given. Considering the low density of non-zero elements in the SCMA codewords, the joint distribution of data symbols is represented as the product of several clique potentials. Then based on the Bethe approximation, the corresponding variational free energy (VFE) is derived. By minimising the VFE, the marginal distributions of symbols are determined. It is shown that the complexity of the proposed receiver only increases linearly with the number of users. Furthermore, a belief damping scheme is employed to improve the performance. Moreover, since the proposed scheme is an iterative one, its convergence is analysed. It is proved that the variance of symbol marginal is guaranteed to converge. For the convergence of its mean, the necessary and sufficient conditions are conceived. Simulation results show that the proposed algorithm approaches the optimal MAP detector's performance despite significant reduction of its complexity.

- *Chapter 5*: In this chapter, low-complexity receiver design for MIMO-SCMA system over frequency selective channels is studied. Auxiliary variables are introduced and the factorisation of joint posterior distribution is represented by a stretched factor graph. Since directly approximating the probability

mass function of discrete symbols as Gaussian leads to performance loss, expectation propagation (EP) is employed to update the messages obtained from channel decoder. Then, using Gaussian approximation of extrinsic information, all messages on factor graph can be parameterised into a Gaussian form, which reduces the computational complexity of message updating significantly. Moreover, considering the proposed BP-EP receiver may fail to converge in the MIMO-SCMA scenario due to the loopy factor graph, the Bethe free energy is convexified and a convergence-guaranteed BP-EP receiver is proposed. Considering the cooperation between users to improve the detection performance, two schemes, namely belief consensus-based and alternative direction method of multipliers (ADMM)-based algorithms are proposed. The simulation results demonstrate the superior performance of the proposed message passing receivers for MIMO-SCMA system and also show the great potentials of using cooperative detection.

- *Chapter 6*: Chapter 6 considers a combined FTN signaling and SCMA system that has even higher spectral efficiency. The system model of an uplink FTN-SCMA system is given first. To tackle the colored noise and ISI induced by FTN signaling and inter user interference induced by SCMA, a novel receiver is designed based on AR model and message passing algorithm that jointly perform channel estimation and detection. Since all messages are represented in Gaussian closed form, the proposed receiver only scales a linear complexity. Moreover, considering a grant free system that the receiver should further detect the active users, a user activity detection, channel estimation and decoding algorithm is developed. With the use of EP approximation of discrete variable indicating user state, a specific factor node is reconstructed, which enables us to still represent all messages in parametric forms. Simulation results show that the combined FTN-SCMA system with the proposed receiver

is capable of increasing the data rate while the BER performance is not affected. Also, in grant free SCMA system, the proposed algorithm is effective to distinguish between the active/inactive users.

- *Chapter 7*: Finally, in the last chapter, the main contributions of this thesis are summarised and some open problems for future work are discussed.

Chapter 2

Joint Channel Estimation and FTN Signaling Detection in Frequency Selective Channels

2.1 Introduction

FTN signaling has higher spectral efficiency than Nqyuist signaling, which makes it attractive in the future communication systems. Nevertheless, the ISI induced by FTN signaling poses challenges on the receiver design. When the channel experiences frequency selective fadings, multipath effect will lead to severer ISI. The interference introduced by FTN signaling and fading channels can be combined to form a composite ISI channel and then the optimal receiver can be designed. However, there are two problems to do so. On one hand, the channel state information may be unknown or not perfectly known. In this circumstance, the channel coefficients have to be estimated. Several papers [64–66] studied channel estimation based on pilot sequence. To obtain accurate CSI, a large number of pilot symbols have to be sent, which is contrary to the FTN system's goal of improving spectral efficiency. On the other hand, in case that the FTN-induced ISI taps are known, combining known and unknown ISI as a new channel and estimate its coefficients may lead to certain performance loss. Since the samples at the receiver side contain information of both data symbols and channel coefficients, a joint channel estimation and decoding algorithm can be developed based on the same received samples. Joint channel estimation and decoding can not only reduce the number of pilots but also enhance the accuracy of channel estimation [67–69]. This chapter considers FTN symbol detection and channel estimation in a frequency selective channel and

develops a low complexity receiver.

Recently, motivated by the heuristic iterative approaches on probabilistic graphical models, several methods are derived for low-complexity iterative receivers based on message passing algorithms, e.g., belief propagation (BP) [70], variational message passing (VMP) [71], and approximate message passing (AMP) [72]. In [73], based on factor graph and BP, a frequency-domain iterative message passing receiver for FTN signaling is proposed in doubly selective channels. The algorithm is evaluated in both perfect and imperfect channel state information (CSI) scenarios. Nevertheless, to the author's best knowledge, joint channel estimation and decoding has not been investigated for FTN signaling. This chapter deals with low-complexity receiver design for FTN signaling in frequency selective fading channels. Based on the Forney-style factor graph, Gaussian message passing (GMP) is employed to update the messages on factor graph. Since the data symbols and channel taps are all unknown variables, conventional message passing algorithm will be infeasible when facing the message updating at the product node. To this end, the factor node is reconstructed, and variational message passing (VMP) is performed. As a result, all messages on factor graph are parameterised by means and variances and the complexity of the proposed algorithm is much lower than the MAP receiver.

2.2 Faster-than-Nyquist Signaling Model

Considering a coded FTN signaling system, at the transmitter side, the information bit sequence $\mathbf{b}$ is encoded to a coded sequence $\mathbf{c}$ and mapped to a length-N data symbol vector $\mathbf{x} = [x_0, ..., x_{N-1}]^T$. The data symbol block passes through the shaping filter $g(t)$, yielding transmitted signal $s(t) = \sum_i g(t - i\tau T_0)x_i$, where $0 \leq \tau \leq 1$ is the FTN packing ratio and T_0 is the symbol period under the Nyquist criterion. Obviously a smaller τ can be chosen to achieve higher data rate at the cost of severer ISI. Note that an FTN symbol is interfered by neighboring symbols

on both sides and the number of ISI taps is infinite. In practice, a sufficiently large number $L_{\text{FTN}} = 2L_f + 1$ can be chosen, then $s(t)$ is given as

$$s(t) = \sum_{i=-L_f}^{L_f} g(t - i\tau T_0)x_i. \tag{2.1}$$

The signal is transmitted over a frequency selective fading channel $h(t)$. For Nyquist signaling, the channel is modeled as L_{nyq} independent taps $\tilde{\mathbf{h}} = [\tilde{h}_{L_{\text{nyq}}-1}, ..., \tilde{h}_0]^T$. However, for FTN signaling, since the symbol period is packed, the number of channel taps becomes greater, i.e, $L = \lfloor L_{\text{nyq}}/\tau \rceil$. According to [74], the channel taps in FTN system can be calculated via interpolation based on the channel taps in Nyquist counterpart. Here DFT is exploited to obtain the fading channel taps for FTN signaling. Assuming $\mathbf{D} \in \mathcal{R}^{L_{\text{nyq}} \times L_{\text{nyq}}}$ and $\mathbf{D}_F \in \mathcal{R}^{L \times L}$ are the DFT matrices, the channel tap $\mathbf{h} = [h_{L-1}, ..., h_l, ..., h_0]^T$ for FTN signaling is given as

$$\mathbf{h} = \mathbf{D}_F^H \begin{bmatrix} \mathbf{D}\tilde{\mathbf{h}} \\ \mathbf{0} \end{bmatrix}. \tag{2.2}$$

Note that the correlations between different channel taps will result in a nondiagonal covariance matrix of $\mathbf{h}$, which is given as

$$\mathbf{V}_h = \mathbf{D}_F^H \begin{bmatrix} \mathbf{D}\mathbf{V}_{\text{nyq}}\mathbf{D}^H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{D}_F,$$

where $\mathbf{V}_{\text{nyq}}$ is the covariance matrix of $\tilde{\mathbf{h}}$.

With the assumption of prefect synchronisation, the received signal can be represented as

$$y(t) = \sum_{l=0}^{L-1} \sum_{i=-L_f}^{L_f} h_l g(t - (i+l)\tau T_0)x_i + n(t), \tag{2.3}$$

where $n(t)$ is white Gaussian noise process with power spectral density N_0. The received continuous-time signal $y(t)$ is matched-filtered by $g^*(t)$ and then sampled

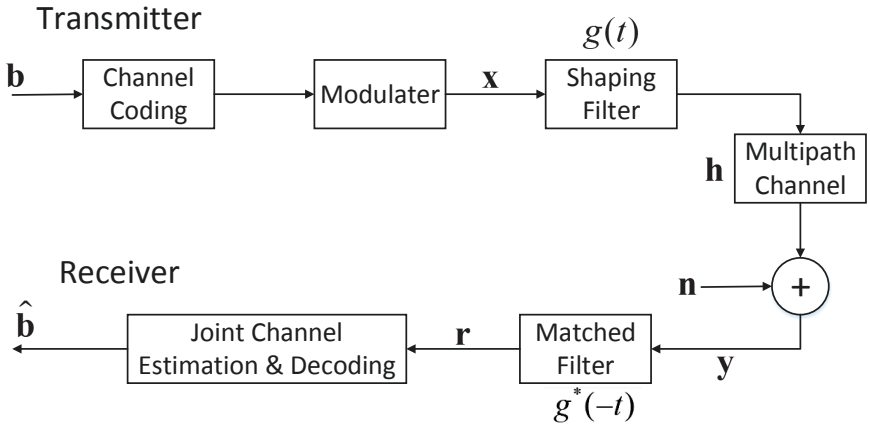


Figure 2.1 : System model for considered FTN signaling system.

with the symbol rate $\frac{1}{\tau T_0}$. At the output of the matched filter, the k-th sample can be expressed as

$$\begin{aligned} r_k &= \sum_{l=0}^{L-1} \sum_{i=-L_f}^{L_f} h_l q(k\tau T_0 - (i+l)\tau T_0) x_i + \xi(k\tau T_0) \\ &= \sum_{l=0}^{L-1} \sum_{i=-L_f}^{L_f} h_l q_{k-l-i} x_i + \xi_k, \end{aligned} \tag{2.4}$$

where $q_{m-n} = \int g(t - m\tau T_0) g^*(t - n\tau T_0) \mathrm{d}t$ and $\xi_k = \int n(t) g^*(t - k\tau T_0) \mathrm{d}t$. Since $g(t)$ is not τT_0-orthogonal, $\{\xi_k\}$ is a colored noise process with autocorrelation function

$$\mathbb{E}[\xi_m \xi_n] = N_0 q_{m-n}. \tag{2.5}$$

With (2.4), the received symbol vector $\mathbf{r} = [r_0, ..., r_{N-1}]^T$ can be written as

$$\mathbf{r} = \mathbf{HQx} + \boldsymbol{\xi}, \tag{2.6}$$

where $\mathbf{x} = [x_0, ..., x_{N-1}]^T$ and $\boldsymbol{\xi} = [\xi_0, ..., \xi_{N-1}]^T$ are the symbol and noise vector; $\mathbf{H}$

and $\mathbf{Q}$ are the matrices with respect to channel tap and FTN ISI tap as

$$\mathbf{H} = \begin{bmatrix} h_0 & & & & & \mathbf{0} \\ h_1 & h_0 & & & & \\ \vdots & & \ddots & & & \\ h_{L-1} & h_{L-2} & \cdots & h_0 & & \\ & \ddots & & & \ddots & \\ \mathbf{0} & & h_{L-1} & \cdots & h_1 & h_0 \end{bmatrix} \tag{2.7}$$

and

$$\mathbf{Q} = \begin{bmatrix} q_0 & q_1 & \cdots & q_{L_f} & & \mathbf{0} \\ \vdots & q_0 & & \ddots & & \\ q_{-L_f} & \cdots & q_0 & \cdots & q_{L_f} & \\ & q_{-L_f} & \cdots & q_0 & \cdots & q_{L_f} \\ & & \ddots & & & \vdots \\ \mathbf{0} & & & q_{-L_f} & \cdots & q_0 \end{bmatrix}. \tag{2.8}$$

The autocorrelation matrix of noise vector $\boldsymbol{\xi}$ is given as $\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^H] = N_0\mathbf{Q}$.

2.3 Message Passing Receiver Design

In this section, a Gaussian message passing based iterative receiver is proposed for joint channel estimation and decoding in FTN signaling system.

2.3.1 Output LLR of Channel Decoder

The receiver performs iterative decoding by exchanging log-likelihood ratio (LLR) between the channel decoder and equalizer. For decoding, the optimal BP decoding algorithm is utilised. Then the output extrinsic LLR of channel decoder can be represented as

$$L^0(c_{n,m}) = \ln \frac{p(c_{n,m} = 0)}{p(c_{n,m} = 1)}, \tag{2.9}$$

where $c_{n,m}$ denotes the mth code bit in the nth subsequence $c_{\mathbf{n}} = [c_{n,1}, ..., c_{n,M}]^T$ with M being the modulation order.

2.3.2 Autoregressive Model of Colored Noise

Due to the correlations between noise samples, conventional MMSE detection approaches suffer from high complexity. Some works neglect the impact of colored noise, which will cause performance loss. To overcome this problem, a Pth-order AR process is employed to approximately model the colored noise [75] , i.e.,

$$\xi_k = \sum_{j=1}^{P} a_j \xi_{k-j} + w_k = \boldsymbol{a}^T \boldsymbol{\xi}_{k-1} + w_k, \tag{2.10}$$

where $\boldsymbol{a} = [a_1, ..., a_P]^T$ is the AR parameters and w_k is the white noise with zero mean and variance σ_w^2, and $\boldsymbol{\xi}_{k-1} = [\xi_{k-1}, ..., \xi_{k-P}]^T$ denotes the correlated noise samples. The autocorrelation parameters $\boldsymbol{a}$ can be obtained from the Yule-Walker equation as

$$N_0 q_k = \begin{cases} N_0 \sum_{j=1}^{P} a_j q_{-j} + \sigma_w^2 & k = 0 \\ N_0 \sum_{j=1}^{P} a_j q_{k-j} & \text{othertwise.} \end{cases} \tag{2.11}$$

2.3.3 Factor Graph Representation

Note that the equation (2.4) can be reformulated as

$$s_k = \mathbf{q}^T \mathbf{x}_k, \tag{2.12}$$

$$r_k = \mathbf{h}^T \mathbf{s}_k + \xi_k, \tag{2.13}$$

where $\mathbf{x}_k = [x_{k-L_f}, ..., x_k, ..., x_{k+L_f}]^T$ and $\mathbf{s}_k = [s_{k-L+1}, ..., s_k]^T$. Moreover, $\mathbf{x}_k$ and $\mathbf{s}_k$ follow state transition model as

$$\mathbf{x}_k = \mathbf{G}\mathbf{x}_{k-1} + \mathbf{f} x_{k+L_f}, \tag{2.14}$$

$$\mathbf{s}_k = \mathbf{G}_1 \mathbf{s}_{k-1} + \mathbf{f}_1^T s_k, \tag{2.15}$$

where the $\mathbf{G} = \begin{bmatrix} \mathbf{0}_{2L_f} & \mathbf{I}_{2L_f} \\ 0 & \mathbf{0}_{2L_f}^T \end{bmatrix}$, $\mathbf{f} = [\mathbf{0}_{2L_f}^T, 1]^T$, $\mathbf{G}_1 = \begin{bmatrix} \mathbf{0}_{L-1} & \mathbf{I}_{L-1} \\ 0 & \mathbf{0}_{L-1}^T \end{bmatrix}$ and $\mathbf{f}_1 = [\mathbf{0}_{L-1}^T, 1]^T$. Similarly, (2.10) can be rewritten as

$$\boldsymbol{\xi}_k = \mathbf{A}\boldsymbol{\xi}_{k-1} + \mathbf{f}_2 w_k, \tag{2.16}$$

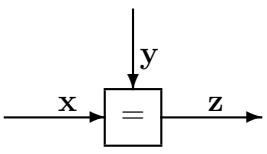
$$\xi_k = \mathbf{f}_2^T \boldsymbol{\xi}_k, \tag{2.17}$$

with $\mathbf{f}_2 = [\mathbf{0}_{P-1}^T, 1]^T$ and $\mathbf{A} = \begin{bmatrix} 0 & \boldsymbol{a}^T \\ \mathbf{0}_{P-1} & \mathbf{I}_{P-1} \end{bmatrix}$.

According to [76], the linear state space model can be represented by a "block diagram" factor graph. Based on (2.12)-(2.17), the corresponding Forney-style factor graph is depicted in Fig. 2.2. On this factor graph, the edges represent variables while the factor nodes denote the local functions. The equality node can be regarded as branching points which allow different factors to share the same variables. Furthermore, a multiplier node $\boxtimes$ is introduced to denote the inner product constraint $\delta(r - \mathbf{h}^T\mathbf{s})$.

2.3.4 Combined BP-EP-VMP Message Passing

GMP is an efficient parametric message passing algorithm in linear Gaussian system, where message on factor graph can be characterised either by the mean vector $\mathbf{m}$ and the covariance matrix $\mathbf{V}$ or by the weight matrix $\mathbf{W} = \mathbf{V}^{-1}$ and the transformed mean $\mathbf{Wm}$.[1] The update rules of GMP have been derived in [76], which is summarised as follows



[1]It may happen frequently that the covariance matrix of a message does not exist due to the singular matrix. Under such circumstances, one may use the transformed means and weight matrices to parameterise the messages [76].

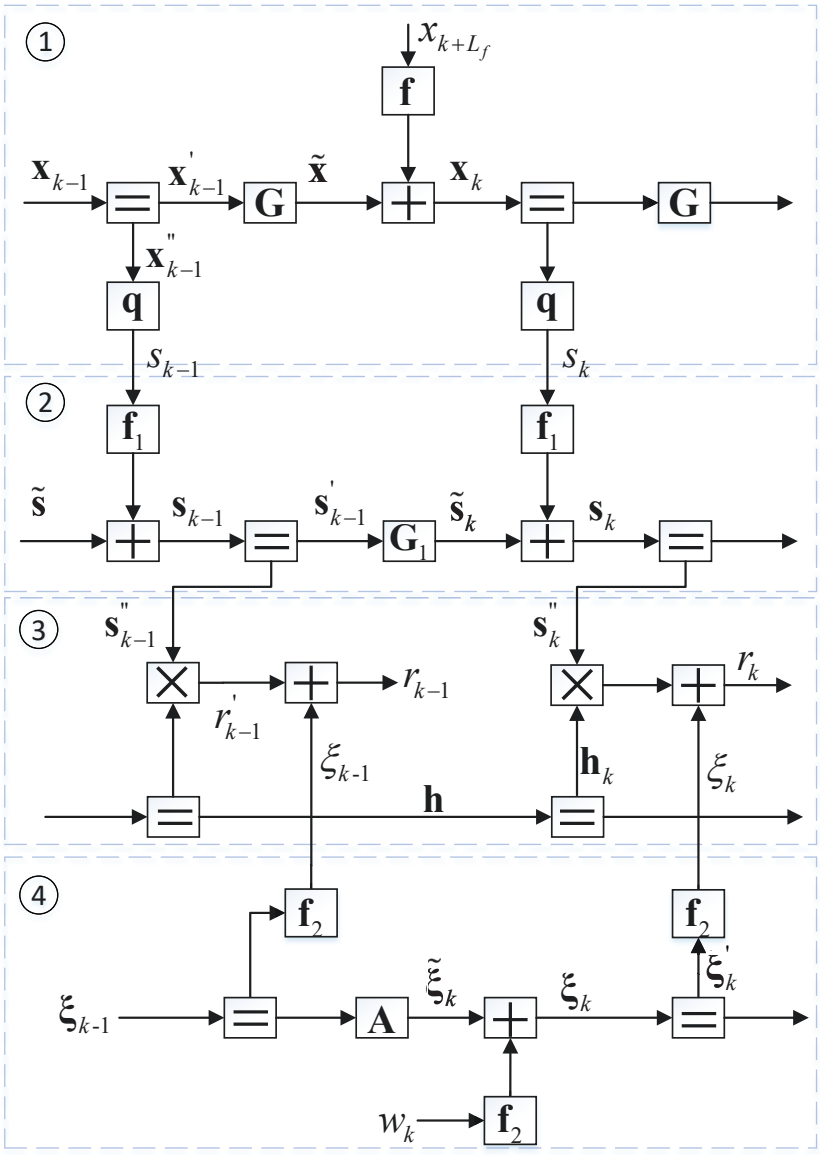


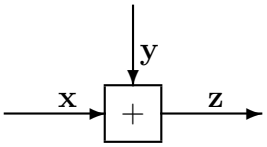
Figure 2.2 : Factor graph representation for joint channel estimation and decoding for FTN system. The subgraphs denoted by ①, ②, ③ and ④ correspond to the FTN equalisation, multipath channel equalisation, channel estimation and colored noise process, respectively.

$$\overrightarrow{\mathbf{W}}_{\mathbf{z}} = \overrightarrow{\mathbf{W}}_{\mathbf{x}} + \overrightarrow{\mathbf{W}}_{\mathbf{y}}$$

$$\overrightarrow{\mathbf{W}}_{\mathbf{z}}\overrightarrow{\mathbf{m}}_{\mathbf{z}} = \overrightarrow{\mathbf{W}}_{\mathbf{x}}\overrightarrow{\mathbf{m}}_{\mathbf{x}} + \overrightarrow{\mathbf{W}}_{\mathbf{y}}\overrightarrow{\mathbf{m}}_{\mathbf{y}}$$

$$\overleftarrow{\mathbf{W}}_{\mathbf{x}} = \overrightarrow{\mathbf{W}}_{\mathbf{y}} + \overleftarrow{\mathbf{W}}_{\mathbf{z}}$$

$$\overleftarrow{\mathbf{W}}_{\mathbf{x}}\overleftarrow{\mathbf{m}}_{\mathbf{x}} = \overrightarrow{\mathbf{W}}_{\mathbf{y}}\overrightarrow{\mathbf{m}}_{\mathbf{y}} + \overleftarrow{\mathbf{W}}_{\mathbf{z}}\overleftarrow{\mathbf{m}}_{\mathbf{z}}$$

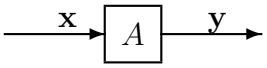


$$\overrightarrow{\mathbf{V}}_{\mathbf{z}} = \overrightarrow{\mathbf{V}}_{\mathbf{x}} + \overrightarrow{\mathbf{V}}_{\mathbf{y}}$$

$$\overrightarrow{\mathbf{m}}_{\mathbf{z}} = \overrightarrow{\mathbf{m}}_{\mathbf{x}} + \overrightarrow{\mathbf{m}}_{\mathbf{y}}$$

$$\overleftarrow{\mathbf{V}}_{\mathbf{x}} = \overleftarrow{\mathbf{V}}_{\mathbf{z}} + \overrightarrow{\mathbf{V}}_{\mathbf{y}}$$

$$\overleftarrow{\mathbf{m}}_{\mathbf{x}} = \overleftarrow{\mathbf{m}}_{\mathbf{z}} - \overrightarrow{\mathbf{m}}_{\mathbf{y}}$$



$$\overrightarrow{\mathbf{m}}_{\mathbf{y}} = \mathbf{A}\,\overrightarrow{\mathbf{m}}_{\mathbf{x}}$$

$$\overrightarrow{\mathbf{V}}_{\mathbf{y}} = \mathbf{A}\,\overrightarrow{\mathbf{V}}_{\mathbf{x}}\,\mathbf{A}^{T}$$

$$\overleftarrow{\mathbf{W}}_{\mathbf{x}} = \mathbf{A}^{T}\,\overleftarrow{\mathbf{W}}_{\mathbf{y}}\,\mathbf{A}$$

$$\overleftarrow{\mathbf{W}}_{\mathbf{x}}\,\overleftarrow{\mathbf{m}}_{\mathbf{x}} = \mathbf{A}^{T}\,\overleftarrow{\mathbf{W}}_{\mathbf{y}}\,\overleftarrow{\mathbf{m}}_{\mathbf{y}}$$

The notations $\overrightarrow{\cdot}$ denotes the message passing along the arrow direction while $\overleftarrow{\cdot}$ denotes the message passing opposite the arrow direction.

Following the GMP rules, most messages on the factor graph can be computed. For ease of exposition, the message updating on the four subgraphs in Fig. 2.2 will be elaborated separately.

■ *Messages Updating for FTN Equalisation (on Subgraph 1):*

Assuming that the parameters $\overrightarrow{\mathbf{W}}_{\mathbf{x}_{k-1}}$ and $\overrightarrow{\mathbf{W}}_{\mathbf{x}_{k-1}}\overrightarrow{\mathbf{m}}_{\mathbf{x}_{k-1}}$ are available, we have

$$\overrightarrow{\mathbf{V}}_{\tilde{\mathbf{x}}_{k-1}} = \mathbf{G}\left(\overleftarrow{\mathbf{W}}_{\mathbf{x}''_{k-1}} + \overrightarrow{\mathbf{W}}_{\mathbf{x}_{k-1}}\right)^{-1}\mathbf{G}^T, \tag{2.18}$$

$$\overrightarrow{\mathbf{W}}_{\tilde{\mathbf{x}}_{k-1}}\overrightarrow{\mathbf{m}}_{\tilde{\mathbf{x}}_{k-1}} = \mathbf{G}\left(\overrightarrow{\mathbf{W}}_{\mathbf{x}_{k-1}}\overrightarrow{\mathbf{m}}_{\mathbf{x}_{k-1}} + \overleftarrow{\mathbf{W}}_{\mathbf{x}''_{k-1}}\overleftarrow{\mathbf{m}}_{\mathbf{x}''_{k-1}}\right), \tag{2.19}$$

where $\overleftarrow{\mathbf{W}}_{\mathbf{x}''_{k-1}}$ and $\overleftarrow{\mathbf{W}}_{\mathbf{x}''_{k-1}}\overleftarrow{\mathbf{m}}_{\mathbf{x}''_{k-1}}$ are given as

$$\overleftarrow{\mathbf{W}}_{\mathbf{x}''_{k-1}} = \frac{\mathbf{q}\mathbf{q}^T}{V_{s_{k-1}}}, \tag{2.20}$$

$$\overleftarrow{\mathbf{W}}_{\mathbf{x}''_{k-1}}\overleftarrow{\mathbf{m}}_{\mathbf{x}''_{k-1}} = \frac{\mathbf{q}m_{s_{k-1}}}{V_{s_{k-1}}}. \tag{2.21}$$

In a similar way, the backward mean and covariance matrix $\overleftarrow{\mathbf{m}}_{\mathbf{x}_k}$ and $\overleftarrow{\mathbf{V}}_{\mathbf{x}_k}$ are obtained as

$$\overleftarrow{\mathbf{V}}_{\mathbf{x}_k} = \left(\overleftarrow{\mathbf{W}}_{\mathbf{x}'_k} + \overleftarrow{\mathbf{W}}_{\mathbf{x}''_k}\right)^{-1}, \tag{2.22}$$

$$\overleftarrow{\mathbf{m}}_{\mathbf{x}_k} = \overleftarrow{\mathbf{V}}_{\mathbf{x}_k}\left(\overleftarrow{\mathbf{W}}_{\mathbf{x}'_k}\overleftarrow{\mathbf{m}}_{\mathbf{x}'_k} + \overleftarrow{\mathbf{W}}_{\mathbf{x}''_k}\overleftarrow{\mathbf{m}}_{\mathbf{x}''_k}\right). \tag{2.23}$$

With (2.18)-(2.23), the outgoing message parameters for x_{k+L_f} are

$$\overleftarrow{m}_{x_{k+L_f}} = \mathbf{f}^T\left(\overleftarrow{\mathbf{m}}_{\mathbf{x}_k} - \overrightarrow{\mathbf{m}}_{\tilde{\mathbf{x}}_{k-1}}\right), \tag{2.24}$$

$$\overleftarrow{V}_{x_{k+L_f}} = \mathbf{f}^T\left(\overrightarrow{\mathbf{V}}_{\tilde{\mathbf{x}}_{k-1}} + \overleftarrow{\mathbf{V}}_{\mathbf{x}_k}\right)\mathbf{f}. \tag{2.25}$$

As the incoming messages are computed from $L^0(c_{n,m})$, they have discrete distributions with respect to the constellation points. In order to employ GMP, EP is used to approximate the incoming messages to be Gaussian [77]. For the kth symbol, the incoming message can be expressed as

$$\overrightarrow{\mu}(x_k) = \sum_{\chi_i \in \mathcal{A}} p_{k,i}\delta(x_k - \chi_i), \tag{2.26}$$

where χ_i is the ith constellation symbol, $\mathcal{A}$ is the set of constellation symbols and $p_{k,i}$ is the probability with respect to χ_i, which is computed from the LLR $L^0(c_{n,m})$.

Then the belief of x_k is obtained as a probability mass function (PMF). Based on EP, matching the first two order moments of the belief yields

$$\tilde{m}_{x_k} = \frac{1}{2\pi \overleftarrow{V}_{x_k}} \sum_{\chi_i \in \mathcal{A}} \chi_i p_{k,i} \exp\left(-\frac{(\chi_i - \overleftarrow{m}_{x_k})^2}{\overleftarrow{V}_{x_k}} \right), \tag{2.27}$$

$$\tilde{V}_{x_k} = \frac{1}{2\pi \overleftarrow{V}_{x_k}} \sum_{\chi_i \in \mathcal{A}} |\chi_i|^2 p_{k,i} \exp\left(-\frac{(\chi_i - \overleftarrow{m}_{x_k})^2}{\overleftarrow{V}_{x_k}} \right) - |m_{x_k}|^2. \tag{2.28}$$

Then the Gaussian approximation to the incoming message can be parameterised as

$$\overrightarrow{m}_{x_k} = \overrightarrow{V}_{x_k} \left(\frac{\tilde{m}_{x_k}}{\tilde{V}_{x_k}} - \frac{\overleftarrow{m}_{x_k}}{\overleftarrow{V}_{x_k}} \right), \tag{2.29}$$

$$\overrightarrow{V}_{x_k} = \left(\tilde{V}_{x_k}^{-1} - \overleftarrow{V}_{x_k}^{-1} \right)^{-1}. \tag{2.30}$$

Consequently, the outgoing messages which are passed to Subgraph 2 read

$$\overrightarrow{m}_{s_k} = \mathbf{q}^T \left(\overrightarrow{\mathbf{W}}_{\mathbf{x}_k} + \overleftarrow{\mathbf{W}}_{\mathbf{x}'_k} \right)^{-1} \left(\overrightarrow{\mathbf{W}}_{\mathbf{x}_k} \left(\overrightarrow{\mathbf{m}}_{\tilde{\mathbf{x}}_{k-1}} + \mathbf{f} \overrightarrow{m}_{x_{k+L_f}} \right) + \overleftarrow{\mathbf{W}}_{\mathbf{x}'_k} \overleftarrow{\mathbf{m}}_{\mathbf{x}'_k} \right), \tag{2.31}$$

$$\overrightarrow{V}_{s_k} = \mathbf{q}^T \left(\overrightarrow{\mathbf{W}}_{\mathbf{x}_k} + \overleftarrow{\mathbf{W}}_{\mathbf{x}'_k} \right)^{-1} \mathbf{q}, \tag{2.32}$$

where $\overrightarrow{\mathbf{W}}_{\mathbf{x}_k} = \left(\overrightarrow{\mathbf{V}}_{\tilde{\mathbf{x}}_{k-1}} + \mathbf{f} \overrightarrow{V}_{x_k} \mathbf{f}^T \right)^{-1}$.

■ *Messages Updating for Multipath Channel Equalisation (on Subgraph 2):*

Similar to (2.24) and (2.25), the backward messages $\overleftarrow{m}_{s_k}$ and $\overleftarrow{V}_{s_k}$ are given by

$$\overleftarrow{m}_{s_k} = \mathbf{f}_1^T \left(\overleftarrow{\mathbf{m}}_{\mathbf{s}_k} - \overrightarrow{\mathbf{m}}_{\tilde{\mathbf{s}}_k} \right), \tag{2.33}$$

$$\overleftarrow{V}_{s_k} = \mathbf{f}_1^T \left(\overrightarrow{\mathbf{V}}_{\tilde{\mathbf{s}}_k} + \overleftarrow{\mathbf{V}}_{\mathbf{s}_k} \right) \mathbf{f}_1, \tag{2.34}$$

where the parameters with respect to $\mathbf{s}_k$ and $\tilde{\mathbf{s}}_k$ have similar form as in (2.18)-(2.23). According to GMP rules, the messages $\overrightarrow{\mathbf{m}}_{\mathbf{s}''_k}$ and $\overrightarrow{\mathbf{V}}_{\mathbf{s}''_k}$ can be expressed as

$$\overrightarrow{\mathbf{m}}_{\mathbf{s}''_k} = \overrightarrow{\mathbf{V}}_{\mathbf{s}''_k} \left(\overrightarrow{\mathbf{W}}_{\mathbf{s}_k} \overrightarrow{\mathbf{m}}_{\mathbf{s}_k} + \overleftarrow{\mathbf{W}}_{\mathbf{s}'_k} \overleftarrow{\mathbf{m}}_{\mathbf{s}'_k} \right), \tag{2.35}$$

$$\overrightarrow{\mathbf{V}}_{\mathbf{s}''_k} = \left(\overrightarrow{\mathbf{W}}_{\mathbf{s}_k} + \overleftarrow{\mathbf{W}}_{\mathbf{s}'_k} \right)^{-1}, \tag{2.36}$$

which are involved in the message computations in Subgraph 3.

■ *Messages Updating for Colored Noise Estimation (on Subgraph 4):*

It is noted that the correlation between colored noise samples does not affect the first-order moment. The means of messages with respect to ξ_k on Subgraph 4 are $\overrightarrow{m}_{\xi_k} = \overleftarrow{m}_{\xi_k} = \mathbb{E}[\xi_k] = 0,\ \forall k$. Therefore only the variances (covariance matrices) of messages need to be calculated. The variance $\overrightarrow{V}_{\xi_k}$ can be obtained as

$$\overrightarrow{V}_{\xi_k} = \mathbf{f}_2^T \overrightarrow{\mathbf{V}}_{\boldsymbol{\xi}_k'} \mathbf{f}_2 = \left[\overrightarrow{\mathbf{V}}_{\boldsymbol{\xi}_k'}\right]_{P,P}, \tag{2.37}$$

where $\overrightarrow{\mathbf{V}}_{\boldsymbol{\xi}_k'}$ is given by

$$\overrightarrow{\mathbf{V}}_{\boldsymbol{\xi}_k'} = \left(\left(\overrightarrow{\mathbf{V}}_{\tilde{\boldsymbol{\xi}}_k} + \sigma_w^2 \mathbf{f}_2 \mathbf{f}_2^T\right)^{-1} + \mathbf{A}^T \overleftarrow{\mathbf{W}}_{\tilde{\boldsymbol{\xi}}_{k+1}} \mathbf{A}\right)^{-1}. \tag{2.38}$$

■ *Messages Updating for Channel Estimation (on Subgraph 3):*

To deal with the inner product of channel vector $\mathbf{h}$ and symbol vector $\mathbf{s}_k$, the messages corresponding to the multiplier node are first considered. With $\overrightarrow{\mathbf{m}}_{\mathbf{s}_k''}$, $\overrightarrow{\mathbf{V}}_{\mathbf{s}_k''}$ and $\overrightarrow{V}_{\xi_k}$ computed on Subgraphs 2 and 4, using BP rule, the message from $\boxtimes$ to $\mathbf{h}_k$ reads

$$\begin{aligned}
\overleftarrow{\mu}(\mathbf{h}_k) &\propto \int \delta(r_k' - \mathbf{h}_k^T \mathbf{s}_k) \overleftarrow{\mu}(\mathbf{s}_k'') \overrightarrow{\mu}(r_k') \, \mathrm{d}\mathbf{s}_k'' \, \mathrm{d}r_k' \\
&\propto \int \exp\left(-(\mathbf{s}_k'' - \overrightarrow{\mathbf{m}}_{\mathbf{s}_k''})^H \overrightarrow{\mathbf{V}}_{\mathbf{s}_k''}^{-1} (\mathbf{s}_k'' - \overrightarrow{\mathbf{m}}_{\mathbf{s}_k''})\right) \exp\left(-\frac{(r_k - \mathbf{h}_k^T \mathbf{s}_k'')^2}{\overrightarrow{V}_{\xi_k}}\right) \mathrm{d}\mathbf{s}_k'' \\
&\propto \exp\left(-\mathbf{h}_k^H \frac{\overrightarrow{\mathbf{m}}_{\mathbf{s}_k''} \overrightarrow{\mathbf{m}}_{\mathbf{s}_k''}^H}{\overrightarrow{V}_{\xi_k} + \mathbf{h}_k^H \overrightarrow{\mathbf{V}}_{\mathbf{s}_k''} \mathbf{h}_k} \mathbf{h}_\mathbf{k} + 2\mathbf{h}_k^H \frac{\overrightarrow{\mathbf{m}}_{\mathbf{s}_k''} r_k}{\overrightarrow{V}_{\xi_k} + \mathbf{h}_k^H \overrightarrow{\mathbf{V}}_{\mathbf{s}_k''} \mathbf{h}_k}\right).
\end{aligned} \tag{2.39}$$

Note that it is difficult to formulate (2.39) to Gaussian. To this end, VMP [78] is used on the multiplier node to derive Gaussian messages. According to VMP update rules, the message $\overleftarrow{\mu}_{\mathbf{h}_k}$ follows

$$\overleftarrow{\mu}(\mathbf{h}_k) \propto \exp\left(\int \ln \delta(r_k' - \mathbf{h}_k^T \mathbf{s}_k'') b(\mathbf{s}_k'') b(r_k') \, \mathrm{d}\mathbf{s}_k'' \, \mathrm{d}r_k'\right), \tag{2.40}$$

where $b(\mathbf{s}_k'')$ and $b(r_k')$ are the beliefs of $\mathbf{s}_k$ and r_k'.

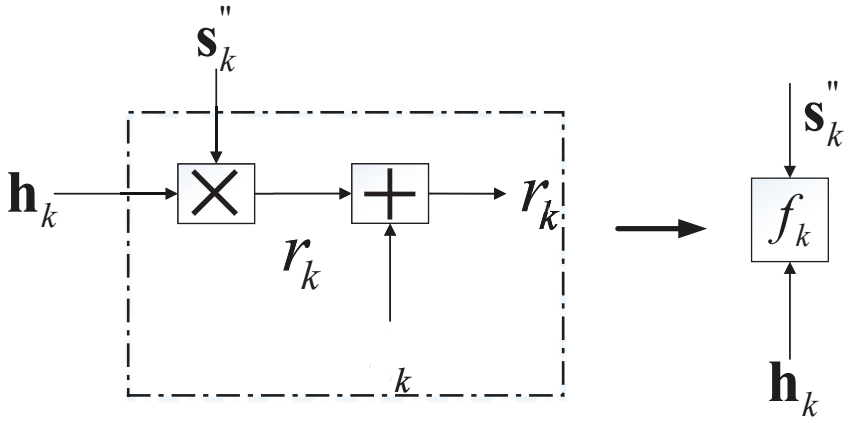


Figure 2.3 : The equivalent "soft node" for multiplier node. The factor f_k denotes the probability density function of r_k conditioned on $\mathbf{s}_k$, $\mathbf{h}_k$ and ξ_k, which can be expressed as $f_k \propto \exp\left(-(r_k - \mathbf{h}_k^T \mathbf{s}_k^{''})^2 / \vec{V}_{\xi_k}\right)$.

Obviously, the logarithm of delta function involved in the integration (2.40) is pathological. To solve this problem, the multiplier node can be grouped with the noisy measurement to form a "soft" factor node f_k [79], as illustrated in Fig. 2.3. Then the message is obtained as

$$\begin{aligned}\overleftarrow{\mu}(\mathbf{h}_k) \propto & \exp\left(-\int \frac{(r_k - \mathbf{h}_k^T \mathbf{s}_k^{''})^2}{\overrightarrow{V}_{\xi_k}} \exp\left(-(\mathbf{s}_k^{''} - \mathbf{m}_{\mathbf{s}_k^{''}})^H \mathbf{V}_{\mathbf{s}_k^{''}}^{-1} (\mathbf{s}_k^{''} - \mathbf{m}_{\mathbf{s}_k^{''}})\right) \mathrm{d}\mathbf{s}_k^{''}\right) \\ \propto & \exp\left(-\mathbf{h}_k^H \frac{\mathbf{V}_{\mathbf{s}_k^{''}} + \mathbf{m}_{\mathbf{s}_k^{''}} \mathbf{m}_{\mathbf{s}_k^{''}}^H}{\overrightarrow{V}_{\xi_k}} \mathbf{h}_k + 2\mathbf{h}_k^H \frac{\mathbf{m}_{\mathbf{s}_k^{''}} r_k}{\overrightarrow{V}_{\xi_k}}\right). \end{aligned} \tag{2.41}$$

Since $\overleftarrow{\mu}(\mathbf{s}_k^{''})$ and $\overrightarrow{\mu}(\mathbf{s}_k^{''})$ have been obtained in Gaussian form, $\mathbf{m}_{\mathbf{s}_k^{''}}$ and $\mathbf{V}_{\mathbf{s}_k^{''}}$ are computed as

$$\mathbf{m}_{\mathbf{s}_k^{''}} = \mathbf{V}_{\mathbf{s}_k^{''}} \left(\overrightarrow{\mathbf{V}}_{\mathbf{s}_k^{''}}^{-1} \overrightarrow{\mathbf{m}}_{\mathbf{s}_k^{''}} + \overleftarrow{\mathbf{V}}_{\mathbf{s}_k^{''}}^{-1} \overleftarrow{\mathbf{m}}_{\mathbf{s}_k^{''}}\right), \tag{2.42}$$

$$\mathbf{V}_{\mathbf{s}_k^{''}} = \left(\overrightarrow{\mathbf{V}}_{\mathbf{s}_k^{''}}^{-1} + \overleftarrow{\mathbf{V}}_{\mathbf{s}_k^{''}}^{-1}\right)^{-1}. \tag{2.43}$$

By noting that the incoming messages to the soft node are beliefs, (2.42) and (2.43) can be regarded as the posterior mean and variance of $\mathbf{s}_k$. Then $\overleftarrow{\mu}(\mathbf{h}_k)$ is calculated

as Gaussian with mean vector and covariance matrix

$$\overleftarrow{\mathbf{V}}_{\mathbf{h}_k} = \overrightarrow{V}_{\xi_k}\left(\mathbf{V}_{\mathbf{s}_k''} + \mathbf{m}_{\mathbf{s}_k''}\mathbf{m}_{\mathbf{s}_k''}^H\right)^{-1}, \tag{2.44}$$

$$\overleftarrow{\mathbf{m}}_{\mathbf{h}_k} = \left(\mathbf{V}_{\mathbf{s}_k''} + \mathbf{m}_{\mathbf{s}_k''}\mathbf{m}_{\mathbf{s}_k''}^H\right)^{-1}\mathbf{m}_{\mathbf{s}_k''} r_k. \tag{2.45}$$

Likewise, $\overleftarrow{\mathbf{V}}_{\mathbf{s}_k''}$ and $\overleftarrow{\mathbf{m}}_{\mathbf{s}_k''}$ can be obtained as

$$\overleftarrow{\mathbf{V}}_{\mathbf{s}_k''} = \overrightarrow{V}_{\xi_k}\left(\mathbf{V}_{\mathbf{h}_k} + \mathbf{m}_{\mathbf{h}_k}\mathbf{m}_{\mathbf{h}_k}^H\right)^{-1}, \tag{2.46}$$

$$\overleftarrow{\mathbf{m}}_{\mathbf{s}_k''} = \left(\mathbf{V}_{\mathbf{h}_k} + \mathbf{m}_{\mathbf{h}_k}\mathbf{m}_{\mathbf{h}_k}^H\right)^{-1}\mathbf{m}_{\mathbf{h}_k} r_k, \tag{2.47}$$

with $\mathbf{m}_{\mathbf{h}_k}$ and $\mathbf{V}_{\mathbf{h}_k}$ computed in a similar way in (2.42) and (2.43).

2.3.5 Computation of Extrinsic LLR

The equalizer calculates the extrinsic LLR based on the soft information $\overleftarrow{m}_{x_k}$ and $\overleftarrow{V}_{x_k}$.

$$\begin{aligned} L^e(c_{n,m}) &= \ln\frac{p(c_{n,m}=0|\mathbf{r})}{p(c_{n,m}=1|\mathbf{r})} - L^0(c_{n,m}) \\ &= \ln\frac{\sum_{\mathbf{d}_{i,m}=0} p(\mathbf{r}|\mathbf{c}_n=\mathbf{d}_i)p(\mathbf{c}_n=\mathbf{d}_i)}{\sum_{\mathbf{d}_{i,m}=0} p(\mathbf{r}|\mathbf{c}_n=\mathbf{d}_i)p(\mathbf{c}_n=\mathbf{d}_i)} - L^0(c_{n,m}), \end{aligned} \tag{2.48}$$

where $\mathbf{d}_i$ is the coded bit sequence corresponding to the constellation symbol χ_i. A concise representation of $L^e(c_{n,m})$ has been derived in [80], which reads

$$L^e(c_{n,m}) = \ln\frac{\sum\limits_{\chi_i\in\mathcal{A}_m^0}\exp\left(-\frac{(\chi_i-\overleftarrow{m}_{x_k})^2}{\overleftarrow{V}_{x_k}}\right)\prod\limits_{m'\neq m} p(c_{n,m'}=s_{i,m'})}{\sum\limits_{\chi_i\in\mathcal{A}_m^1}\exp\left(-\frac{(\chi_i-\overleftarrow{m}_{x_k})^2}{\overleftarrow{V}_{x_k}}\right)\prod\limits_{m'\neq m} p(c_{n,m'}=s_{i,m'})}, \tag{2.49}$$

where $\mathcal{A}_m^0$ and $\mathcal{A}_m^1$ denote the subset of $\mathcal{A}$ whose label in position m has the value 0 or 1. Then the LLRs $\{L^e(c_{n,m})\}$ are fed to the channel decoder. After decoding, the decoder outputs the extrinsic LLR and turns to the next iteration of equalisation. The details of the proposed BP-EP-VMP algorithm are summarised in Algorithm 1.

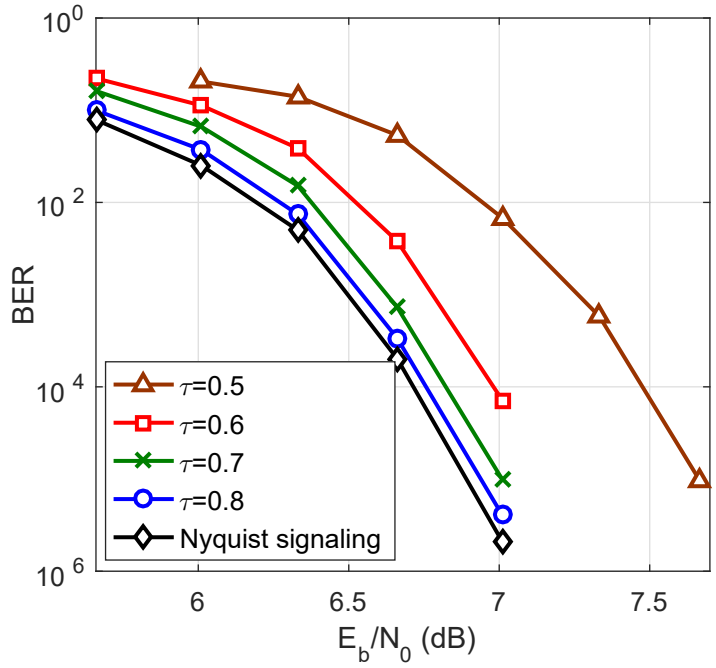
2.3.6 Complexity Analysis

The complexity of the proposed algorithm is dominated by the matrix inversion operations in (2.18), (2.22), (2.36), (2.38), (2.43) and (2.44). For a non-sparse $K \times K$ matrix, the complexity for calculating its inverse is $\mathcal{O}(K^3)$. Here $\mathcal{O}$ denotes the order of time complexity. Taking (2.18) as an example, the total computational complexity is $\mathcal{O}(NL_{\text{FTN}}^3)$ for a length-N symbol block. Then the total computational complexity of the proposed algorithm is $\mathcal{O}(N(L_{\text{FTN}}^3 + L^3 + P^3))$, where L_{FTN} is the length of FTN-induced ISI considered at the receiver, L is the channel length and P is the order of AR model used to approximate the colored noise.

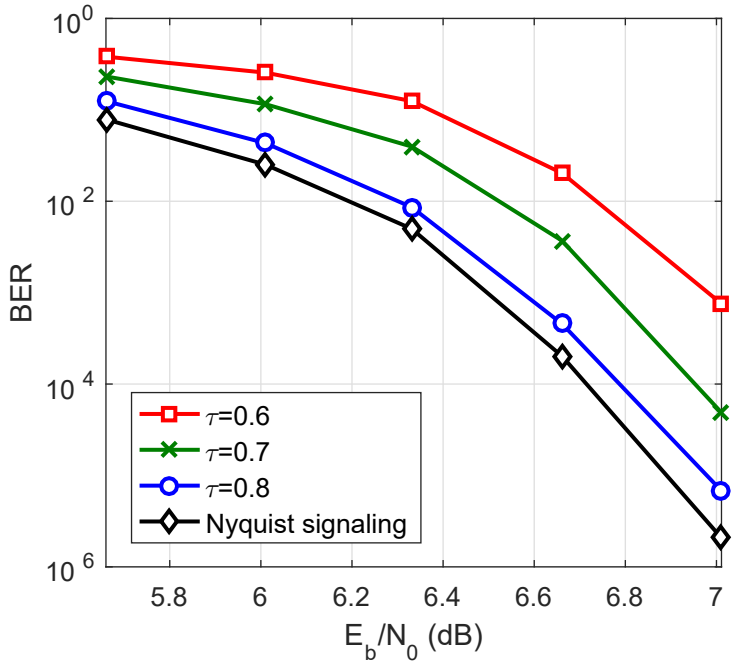
2.4 Simulation Results

In the simulations, a 5/7-rate LDPC code is considered with variable and check node degree distributions being $v(X) = 0.0005 + 0.2852X + 0.2857X^2 + 0.4286X^3$, $c(X) = 0.0017X^9 + 0.9983X^{10}$ [81]. The encoded bits are interleaved and mapped to a sequence of QPSK symbols. The number of transmitted symbols is $N = 2048$. The sequence of symbols passes through a root raised-cosine shaping filter with a roll-off factor $\alpha = 0.4$. The carrier frequency $f_0 = 2$GHz and the symbol period $T = 0.2\mu$s. The frequency selective fading channel for Nyquist signaling is assumed to have $L = 20$ taps and the coefficients $\{\tilde{h}_l\}$ are independently generated according to the distribution $\tilde{h}_l \sim \mathcal{N}(0, q^l)$. Then the channel taps for FTN signaling can be obtained by interpolation using DFT matrices. The normalised power delay profile is $q^l = \frac{\exp(-0.05l)}{\sum q^l}$. The number of iterations is set to $I = 10$ and the maximum number of BP decoding iterations is 50. The number of ISI taps due to FTN considered by the receiver is $L_{\text{FTN}} = 11$, unless otherwise specified. All simulation results are averaged over 1000 independent Monte Carlo trails.

The impact of packing factor τ is first evaluated. As shown in Fig. 2.4(a), BER performance versus the signal-to-noise ratio (SNR) of the proposed BP-EP-VMP



(a) $\alpha = 0.4$



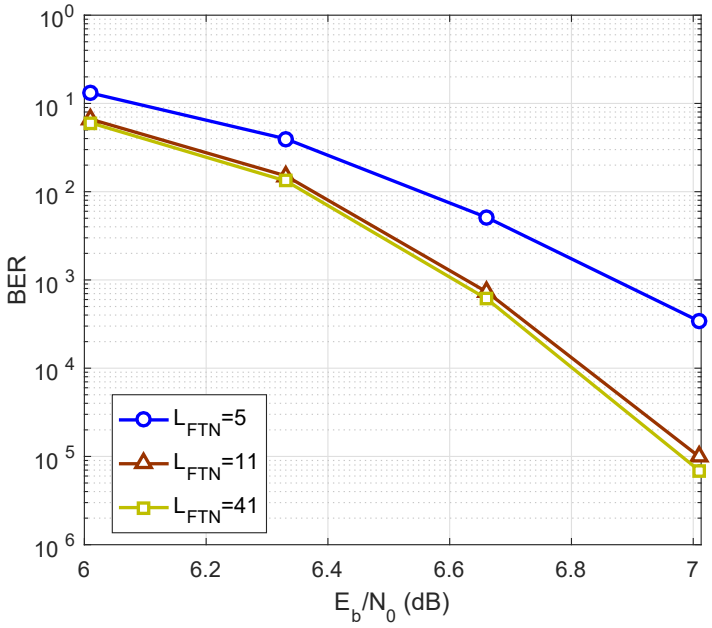
(b) $\alpha = 0.05$

Figure 2.4 : BER performance of the proposed algorithm for FTN system with different packing factor τ. The roll-off factor $\alpha = 0.4$ and $\alpha = 0.05$, respectively.

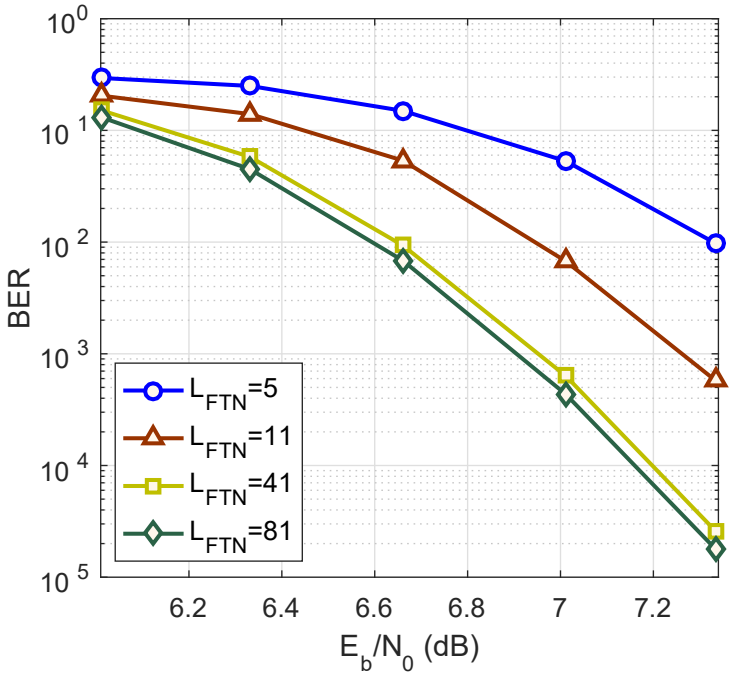
algorithm with various τ are plotted. The performance of Nyquist signaling over the same channel is also included as a benchmark. It is observed that the proposed FTN receiver can attain the BER performance of the Nyquist signaling when the packing factor $\tau \geq 0.7$. Therefore, up to 40% of transmission rate can be increased with the same bandwidth by employing FTN signaling. Even for $\tau = 0.6$, the performance gap is less than 0.2dB, while the transmission rate in this case can be increased by more than 65%. In Fig. 2.4(b), the BER performance with roll-off factor $\alpha = 0.05$ is evaluated. Since smaller α will lead to stronger ISI, it is seen that, compared with the Nyquist counterpart, the performance gap for $\tau = 0.6$ becomes about 0.5dB. Nevertheless, it is able to improve transmission rate up to 25% by employing FTN signaling with $\tau = 0.8$.

The complexity of the proposed algorithm depends on L_{FTN}, i.e., the length of FTN-induced ISI considered at receiver. In Fig. 2.5(a), BER performances with different $L_{\text{FTN}} = \{5, 11, 41\}$ are illustrated, where the roll-off factor $\alpha = 0.4$ and $\tau = 0.7$. It is seen that, $L_{\text{FTN}} = 5$ suffers from significant performance degradation due to the underestimation of ISI induced by FTN. The increase of L_{FTN} helps to improve the BER performance, and the gain becomes marginal when $L_{\text{FTN}} \geq 11$. Therefore, the length of ISI induced by FTN signaling can be safely approximated by $L_{\text{FTN}} = 11$ in this case. Further consider a stronger packing scenario with $\tau = 0.5$ and the BER performance is illustrated in Fig. 2.5(b). It is seen that $L_{\text{FTN}} = 11$ is not long enough to approximate the length of FTN-induced ISI and about 0.3dB performance loss can be observed compared to the $L_{\text{FTN}} = 41$ case. Therefore, In practice, BER performance and computational complexity can be compromised by selecting a proper value of L_{FTN}.

In Fig. 2.6, BER performance of the proposed BP-EP-VMP algorithm is compared with other methods. Since there is no existing work on this topic, two Bayesian estimators are extended to the FTN signaling over frequency selective channels,



(a) $\tau = 0.7$



(b) $\tau = 0.5$

Figure 2.5 : Impact of L_{FTN} on BER performance. The roll-off factor $\alpha = 0.4$, $\tau = 0.7$ and $\tau = 0.5$, respectively.

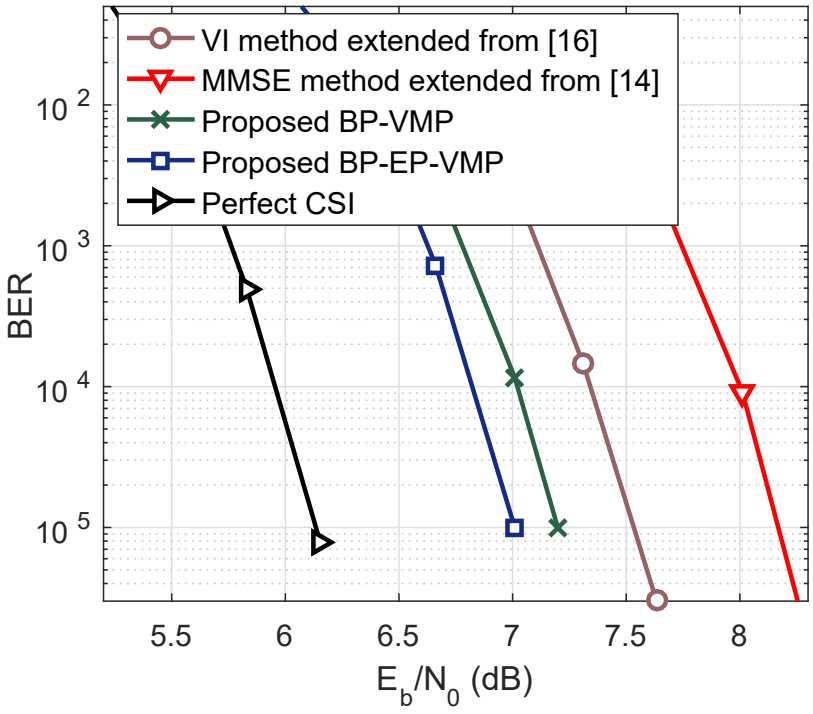


Figure 2.6 : BER performance of different algorithms for considered FTN signaling system, with $\tau = 0.7$, $\alpha = 0.4$.

namely, the MMSE equalizer [67] and the variational inference (VI) method [69]. The BER performance with perfect channel information is also included as a reference. The MMSE equalizer can only treat the ISI caused by FTN signaling and the fading channel as a composite ISI channel, which leads to increased number of channel coefficients to be estimated. The VI method suffers from performance degradation due to its assumption that data symbols are conditional independent. The complexities of the MMSE equalizer and the VI method are $\mathcal{O}(N^3)$ and $\mathcal{O}(N(L_{\text{FTN}}^2 + L^2 + P^2))$, respectively. It is seen that the proposed BP-EP-VMP algorithm outperforms the other methods. When BP instead of EP is employed to update the messages from output of channel decoder to the equalizer, denoted as "BP-VMP" in In Fig. 2.6, about 0.2dB performance loss can be observed, which demonstrates the superior performance by employing EP.

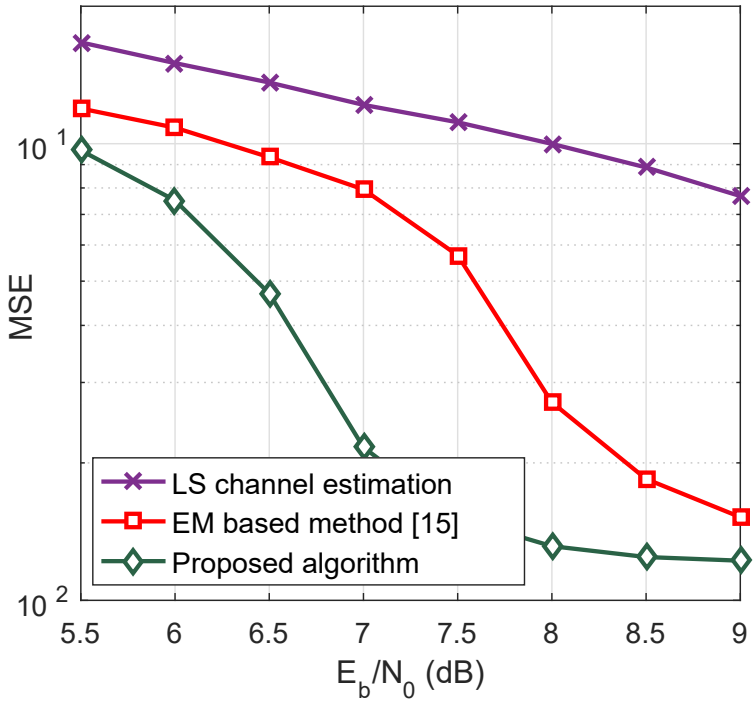


Figure 2.7 : MSE of channel estimation of the proposed algorithm, with $\tau = 0.7$, $\alpha = 0.4$.

The mean squared error (MSE) of channel estimation of the proposed BP-EP-VMP algorithm is evaluated in Fig. 2.7. For comparison, the performances of least square (LS) channel estimation [65] and the expectation-maximisation (EM)-based method [68] are also included. The number of channel taps is 14 in $\tau = 0.7$ scenario. It is seen that the proposed BP-EP-VMP algorithm and the EM-based method significantly outperform the LS channel estimation that only uses the limited amount of pilot symbols. Moreover, the proposed BP-EP-VMP algorithm is superior to the EM-based method in [68], since the latter only provides hard channel estimation to the equalizer. Simulation results corroborate the benefits of the proposed joint channel estimation and decoding scheme.

2.5 Conclusions

In this chapter, a low-complexity FTN receiver is proposed to perform joint channel estimation and decoding in frequency selective fading channels. The ISI imposed by FTN signaling and that by the unknown frequency selective channel are considered separately, which enable us to fully exploit the known structure of the FTN-induced ISI. Colored noise due to the faster sampling rate of FTN signaling is approximated by AR process to avoid using whitening filter. A Forney style factor graph, which consists of four subgraphs, is constructed to represent the FTN system. It is shown that using BP on the factor graph directly is infeasible, since the messages corresponding to the inner product of FTN symbol vector and the channel coefficient vector cannot be updated efficiently. VMP is performed on an equivalent "soft node" to tackle this problem. Moreover, EP is employed to efficiently convert the messages corresponding to FTN symbols obtained from channel decoder to Gaussian distribution. It is shown that, since the proposed hybrid BP-EP-VMP algorithm enables effective Gaussian message approximation, the complexity only increases linearly with the block length N. Simulation results show that the superior performance of the proposed algorithm compared with the existing methods in FTN system. Compared with the Nyquist counterpart, FTN signaling with the proposed algorithm is able to increase the transmission rate over 40% in frequency selective fading channels, with negligible BER performance loss.

Algorithm 1 The Proposed Hybrid BP-EP-VMP Approach to Joint Channel Estimation and Decoding for FTN Signaling over Frequency Selective Channels

1: **Initialisation:**

The output LLRs of channel decoder are initialised as $L^0(c_{m,n}) = 0$, i.e., $\overrightarrow{m}^0_{x_k} = 0$ and $\overrightarrow{V}^0_{x_k} = +\infty$.

The initial estimation of channel coefficients are obtained using 20 pilot symbols based on least square method.(In each turbo equalisation, pilot symbols are also used to estimate the channel taps.) Then, the prior of $\mathbf{h}$ can be expressed as

$$p(\mathbf{h}) \propto \exp\left((\mathbf{h} - \overrightarrow{\mathbf{m}}^0_{\mathbf{h}})^H \overrightarrow{\mathbf{W}}^0_{\mathbf{h}} (\mathbf{h} - \overrightarrow{\mathbf{m}}^0_{\mathbf{h}})\right),$$

where $\overrightarrow{\mathbf{m}}^0_{\mathbf{h}}$ is the measured CSI and $\overrightarrow{\mathbf{W}}^0_{\mathbf{h}}$ is diagonal with the entries being $\left[\overrightarrow{\mathbf{W}}^0_{\mathbf{h}}\right]_{ii} = \sigma^2_{h_i}$. $\overrightarrow{m}^0_{x_k} = 0$, $\overrightarrow{V}^0_{x_k}$, $\overrightarrow{\mathbf{m}}^0_{\mathbf{h}}$ and $\overrightarrow{\mathbf{W}}^0_{\mathbf{h}}$ can be regarded as the prior information.

2: **for** Iter=1:I **do**

3: Calculate the messages from subgraph 1 to subgraph 2 according to (2.31) and (2.32);

4: Calculate the messages on subgraph 2 according to (2.33)-(2.36);

5: Calculate the messages on subgraph 3 according to (2.37) and (2.38);

6: Calculate the messages related to the "multiplier node" using (2.44)-(2.47);

7: Convert the outgoing messages to LLR based on (2.49) and feed them to channel decoder;

8: Perform BP channel decoding algorithm;

9: Calculate the incoming messages using EP as in (2.29) and (2.30);

10: **end for**

Chapter 3

Low Complexity Receiver Design for FTN Signaling in Doubly Selective Channels

3.1 Introduction

In chapter 2, the iterative receiver design for FTN signaling over frequency selective fading channels is studied. When the channel state information is known, the FDE-MMSE based algorithm [53] can detect the data symbols in a low complexity fashion. However, in high mobility environments, the channel coefficients vary dramatically with respect to time and the time selective fading caused by Doppler shift is unavoidable. This means the receiver should take interferences induced by frequency selectivity as well as time selectivity into account, which is the data detection problem in doubly selective channels (DSCs). Since the number of channel coefficients become very large in DSCs, conventional MAP estimation suffers from huge complexity. To this end, several works studied the low compleixty receiver design in DSCs based on shortening the number of interferences [82, 83], basis expansion model [84] and compressive sensing [85, 86]. Although the receiver complexity was reduced, it still increases exponentially. In [87], the authors proposed a linear MMSE (LMMSE) based method for slow variant DSCs. Nevertheless, all above researches focused on Nyquist signaling and the reception of FTN signaling over DSCs has not been considered. For FTN signaling, the FDE-MMSE algorithm [53] can be extended to a DSC case, but the complexity is very high due to the inter carrier interference. Therefore, it is of importance to investigate low complexity FDE method for FTN signaling in DSCs.

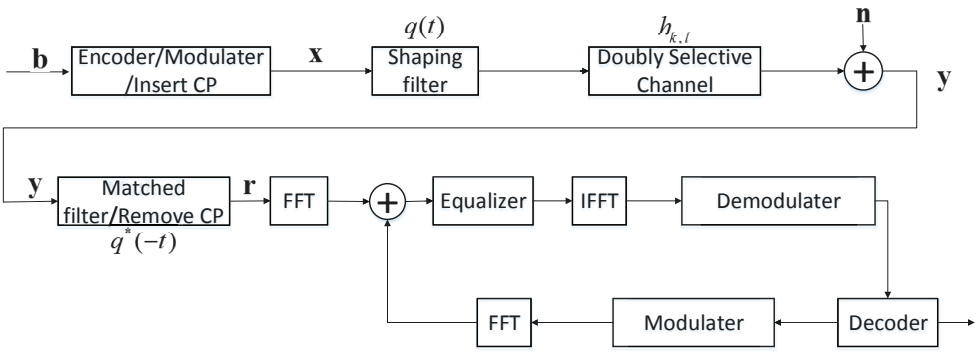


Figure 3.1 : System Model for FTN signaling in DSCs.

This chapter will study the FDE for FTN signaling in DSCs. Firstly, the system model of FTN signaling in frequency domain is introduced and then the extension of the conventional FDE-MMSE is presented. Next, the posterior distribution is written as the product of potential functions and two algorithms based on mean field (MF) [88] and Bethe [89] approximations, respectively, are proposed via minimising the corresponding variational free energy (VFE). In another way, by approximately considering the interference only from neighboring frequency-domain symbols, it is able to reformulate the problem of FTN signaling detection in DSCs into a linear state-space model. Then, Gaussian message passing (GMP) is performed on a vector form factor graph to design a low complexity FTN receiver. Considering the imperfect channel information scenario, the channel uncertainty is taken into account on the factor graph and robust message passing receiver for FTN signaling is derived. Finally, the effectiveness of both algorithms are verified via Monte Carlo simulations.

3.2 System Model

The system model is given in Fig. 3.1. A sequence of information bits $\mathbf{b}$ which is coded and mapped to a sequence of phase-shift keying (PSK)/ quadrature amplitude

modulation (QAM) symbols $\mathbf{x} = [x_0, ..., x_{N-1}]^T$. After inserting $2M$ symbols of cyclic prefix (CP), the symbol vector is passed through a shaping filter $q(t)$ with symbol interval $T_0 = \tau T$, where $0 < \tau \leq 1$ is the packing ratio of FTN signaling. Note that $\tau = 1$ corresponds to Nyquist signaling. Obviously, the transmission rate is increased by $1/\tau$ times in FTN. Then, the signal is transmitted over a DSC with impulse response at the kth symbol interval $[h_{k,0}, ..., h_{k,L-1}]$.

With the assumption of perfect synchronisation, the received signal after matched filtering is sampled at time instant $k\tau T$. After removing the CP, the kth sample can be expressed as [53, 90]

$$r_k = \sum_{l=0}^{L-1} h_{k,l} \sum_{i=-M}^{M} q_i x_{k-l-i} + \gamma_k, \tag{3.1}$$

with

$$q_{k-m} = \int g(t - k\tau T) g^*(t - m\tau T) \mathrm{d}t, \tag{3.2}$$

$$\eta_k = \int n(t) g^*(t - k\tau T) \mathrm{d}t, \tag{3.3}$$

where $n(t)$ is an additive white Gaussian noise process with zero mean and variance N_0. Since $q(t)$ is non-orthogonal with period τT, the noise η_k is colored with autocorrelation function given as

$$\mathbb{E}[\eta_k \eta_m] = N_0 q_{k-m}. \tag{3.4}$$

The received signal vector $\mathbf{r} \triangleq [r_0, ..., r_{N-1}]^T$ can be rewritten in a matrix form as [91, 92]

$$\mathbf{r} = \mathbf{HQx} + \boldsymbol{\eta}, \tag{3.5}$$

where $\mathbf{H}$ and $\mathbf{Q}$ are $N \times N$ matrices with the kth row being $[\mathbf{0}_k, h_{k,0}, ..., h_{k,L-1}, \mathbf{0}_{N-k-L}]$ and $[\mathbf{0}_k, q_{-M}, ..., q_M, \mathbf{0}_{N-k-2M}]$, respectively; $\boldsymbol{\eta} \triangleq [\eta_0, ..., \eta_{N-1}]^T$ is the colored noise vector with autocorrelation matrix $\mathbb{E}[\boldsymbol{\eta}\boldsymbol{\eta}^H] = N_0\mathbf{Q}$.

The received signal $\mathbf{r}$ can be transformed to its frequency-domain counterpart as

$$\mathbf{r}_f = \mathbf{FHQx} + \boldsymbol{\eta}_f, \tag{3.6}$$

where $\mathbf{F}$ is the $N \times N$ normalised discrete Fourier transform (DFT) matrix, i.e., $\mathbf{F}_{m,n} = \frac{1}{\sqrt{N}} \exp(-2\pi jmn)$. Note that $\mathbf{Q}$ is a circulant matrix with eigenvalue decomposition $\mathbf{Q} = \mathbf{F}^H \mathbf{\Lambda F}$, where $\mathbf{\Lambda} = \text{diag}\{\lambda_0, .., \lambda_{N-1}\}$ and λ_i is the ith eigenvalue of $\mathbf{Q}$. Then $\mathbf{r}_f$ can be formulated as

$$\mathbf{r}_f = \mathbf{C\Lambda x}_f + \boldsymbol{\eta}_f, \tag{3.7}$$

where $\mathbf{x}_f = \mathbf{Fx} = [x_{f,0}, ..., x_{f,N-1}]^T$, $\mathbf{C} = \mathbf{FHF}^H$ are the frequency-domain symbol vector and channel matrix, respectively, $\boldsymbol{\eta}_f = \mathbf{F}\boldsymbol{\eta}$ is the frequency-domain noise vector with autocorrelation matrix

$$\mathbb{E}[\boldsymbol{\gamma}_f {\boldsymbol{\gamma}_f}^H] = \mathbb{E}[\mathbf{F}\boldsymbol{\gamma}\boldsymbol{\gamma}^H \mathbf{F}^H] = N_0 \mathbf{FQF}^H = N_0 \mathbf{\Lambda}. \tag{3.8}$$

3.3 FDE-MMSE based Algorithm

The frequency-domain symbol vector $\mathbf{x}_f$ in (3.7) can be detected using iterative MMSE equalizer, with weights $\mathbf{W}$ given by

$$\mathbf{W} = \mathbf{\Lambda}^H \mathbf{C}^H (\mathbf{C\Lambda\Lambda}^H \mathbf{C}^H + N_0 \mathbf{\Lambda})^{-1}. \tag{3.9}$$

And the estimate of frequency domain symbols $\mathbf{x}_f$ is given by

$$\hat{\mathbf{x}}_f = \mathbf{W}\mathbf{r}_f. \tag{3.10}$$

Then employing inverse DFT (IDFT) can provide the estimates of data symbols.

The computational complexity of MMSE equalizer mainly depends on the matrix inversion operation. For frequency selective fading channels, since $\mathbf{C}$ is diagonal, the complexity of matrix inversion becomes $\mathcal{O}(N)$ [90]. However, for DSCs considered in this letter, $\mathbf{C}$ is no longer diagonal and the complexity becomes $\mathcal{O}(N^3)$. A brute

force solution is to approximate $\mathbf{C}$ by its diagonal entries. Accordingly, $\mathbf{W}$ becomes diagonal with the (i,i)th element given by

$$\mathbf{W}_{i,i} = \frac{\mathbf{C}_{i,i}\lambda_i}{|\mathbf{C}_{i,i}\lambda_i|^2 + N_0\lambda_i}. \tag{3.11}$$

However, this approximation suffers significant performance loss since interference between frequency-domain symbols are underestimated.

3.4 Variational Inference-based FDE for FTN Signaling in DSCs

3.4.1 Probabilistic Model

According to the Bayesian rule, the *a posteriori* distribution of frequency-domain symbol vector $\mathbf{x}_f$ reads

$$p(\mathbf{x}_f|\mathbf{r}_f) \propto p(\mathbf{r}_f|\mathbf{x}_f)p(\mathbf{x}_f), \tag{3.12}$$

where $p(\mathbf{r}_f|\mathbf{x}_f)$ is the likelihood function expressed as

$$p(\mathbf{r}_f|\mathbf{x}_f) \propto \exp(-(\mathbf{r}_f - \mathbf{C}\mathbf{\Lambda}\mathbf{x}_f)^H(N_0\mathbf{\Lambda})^{-1}(\mathbf{r}_f - \mathbf{C}\mathbf{\Lambda}\mathbf{x}_f)), \tag{3.13}$$

$p(\mathbf{x}_f)$ is the *a priori* distribution of $\mathbf{x}_f$ with mean vector $\mathbf{m}^0_{\mathbf{x}_f} = \mathbf{F}\mathbf{m}^0_{\mathbf{x}}$ and covariance matrix $\mathbf{V}^0_{\mathbf{x}_f} = \mathbf{F}\mathbf{V}^0_{\mathbf{x}}\mathbf{F}^H$, and $\mathbf{m}^0_{\mathbf{x}} = [m^0_{x_0}, ..., m^0_{x_{N-1}}]^T$ and $\mathbf{V}^0_{\mathbf{x}} = \text{diag}\{V^0_{x_0}, ..., V^0_{x_{N-1}}\}$ are the mean vector and covariance matrix of time-domain symbol vector $\mathbf{x}$, which are calculated based on the extrinsic LLR from output of the channel decoder at the previous iteration. To simplify the calculations, the approximation [93] is employed,

$$\mathbf{V}^0_{\mathbf{x}} \simeq a\mathbf{I}, \text{ with } a = \frac{\sum_{i=0}^{N-1} V_{x_i^0}}{N}. \tag{3.14}$$

Consequently, $\mathbf{V}^0_{\mathbf{x}_f} = a\mathbf{I}$. Since different samples in $\boldsymbol{\eta}_f$ are independent, (3.12) can be rewritten as follows,

$$p(\mathbf{x}_f|\mathbf{r}_f) \propto \prod_i \underbrace{p(x_{f,i})\exp(-\mathcal{R}\{d_{ii}\}|x_{f,i}|^2 + \mathcal{R}\{d_i x^*_{f,i}\})}_{\varphi_{ii}(x_{f,i})} \prod_{i,j} \underbrace{\exp(-\mathcal{R}\{d_{ij}x^*_{f,i}x_{f,j}\})}_{\varphi_{ij}(x_{f,i},x_{f,j})}, \tag{3.15}$$

where $d_{ij} = \mathbf{C}_{:,i}^H (N_0 \mathbf{\Lambda})^{-1} \mathbf{C}_{:,j}$ and $d_i = \frac{\mathbf{C}_{:,i}^H \mathbf{r}_f}{N_0 \lambda_i}$. Without loss of generality, let $\varphi_{ij} = \varphi_{ji}$ to simplify the notations hereafter.

3.4.2 Variational Inference Method

The maximum *a posteriori* probability (MAP) detection of $x_{f,i}$ can be obtained by marginalising $p(\mathbf{x}_f|\mathbf{r}_f)$ in (3.15), i.e., $p(x_{f,i}|\mathbf{r}_f) = \sum_{\mathbf{x}_f \backslash x_{f,i}} p(\mathbf{x}_f|\mathbf{r}_f)$, which involves huge computational complexity. The variational inference method is employed to approximate $p(\mathbf{x}_f|\mathbf{r}_f)$ by a proper distribution $q(\mathbf{x}_f)$ which is easy to be marginalised. The approximation can be achieved by minimising the variational free energy (VFE) [94] formulated as

$$\mathcal{F} = \int q(\mathbf{x}_f) \ln \frac{q(\mathbf{x}_f)}{p(\mathbf{x}_f|\mathbf{r}_f)} \mathrm{d}\mathbf{x}_f + \mathrm{C}_0, \tag{3.16}$$

where C_0 is a constant value.

The MF approximation $q(\mathbf{x}_f) = \prod_i q(x_{f,i})$ is first considered, where all variables in $\mathbf{x}_f$ are assumed to be conditional independent. Minimising MF VFE yields the approximate marginal $q(x_{f,i})$ as (See Appendix. A for details)

$$q(x_{f,i}) \propto \varphi_{ii}(x_{f,i}) \prod_{j=0, j\neq i}^{N-1} \exp\left(\int \ln \varphi_{ij}(x_{f,i}, x_{f,j}) q(x_{f,j}) \mathrm{d}x_{f,j} \right). \tag{3.17}$$

With the Gaussian assumption $q(x_{f,j}) \propto \mathcal{CN}(x_{f,j}, m_{x_{f,j}}, v_{x_{f,j}})$, $q(x_{f,i})$ is characterised by its mean $m_{x_{f,i}}$ and variance $v_{x_{f,i}}$. Then, the mean and variance for the output extrinsic information are given by $v_{x_{f,i}}^e = (1/v_{x_{f,i}} - 1/v_{x_{f,i}}^0)^{-1}$, $m_{x_{f,i}}^e = v_{x_{f,i}}^e (m_{x_{f,i}}/v_{x_{f,i}} - m_{x_{f,i}}^0/v_{x_{f,i}}^0)$, which can be expressed as

$$m_{x_{f,i}}^e = \frac{d_i - \prod_{j=0, j\neq i}^{N-1} d_{ij} m_{x_{f,j}}}{\mathcal{R}\{d_{ii}\}}, \tag{3.18}$$

$$v_{x_{f,i}}^e = \frac{1}{\mathcal{R}\{d_{ii}\}}, \tag{3.19}$$

with $m_{x_{f,j}} = \mathbb{E}_{q(x_{f,j})}(x_{f,j})$. It is observed from (3.19) that with MF approximation, the variance of $x_{f,i}$ is irrelevant to the variances of other symbols. In other words,

the independence assumption in MF approximation leads to the underestimation of variance, which may result in performance loss.

To improve the performance of MF approximation, the Bethe approximation is employed which takes into account the conditional dependencies between frequency-domain symbols. The Bethe approximation is given by

$$q(\mathbf{x}_f) = \prod_i q(x_{f,i}) \prod_{i,j} \frac{q(x_{f,i}, x_{f,j})}{q(x_{f,i})q(x_{f,j})}. \tag{3.20}$$

Minimising Bethe VFE yields (See Appendix. B for details)

$$q(x_{f,i}) \propto \varphi_{ii}(x_{f,i}) \prod_{j=0, j\neq i}^{N-1} \int \varphi_{ij}(x_{f,i}, x_{f,j}) \tilde{q}_{\backslash i}(x_{f,j}) \mathrm{d}x_{f,j}, \tag{3.21}$$

where $\tilde{q}_{\backslash i}(x_{f,j}) = q(x_{f,j}) / \int \varphi_{ij}(x_{f,i}, x_{f,j}) \tilde{q}_{\backslash j}(x_{f,i}) \mathrm{d}x_{f,i}$. Assuming that $\tilde{q}_{\backslash i}(x_{f,j})$ is available in Gaussian form $\mathcal{CN}(x_{f,j}, \tilde{m}_{\backslash i, x_{f,j}}, \tilde{v}_{\backslash i, x_{f,j}})$, the mean and variance for the extrinsic information can be determined as

$$m^e_{x_{f,i}} = \frac{d_i - \prod_{j=0, j\neq i}^{N-1} d_{ij} \tilde{m}_{\backslash i, x_{f,j}}}{\mathcal{R}\{d_{ii}\} + \tilde{v}_{\backslash i, x_{f,j}} |d_{ij}|^2}, \tag{3.22}$$

$$v^e_{x_{f,i}} = \frac{1}{\mathcal{R}\{d_{ii}\} + \tilde{v}_{\backslash i, x_{f,j}} |d_{ij}|^2}. \tag{3.23}$$

The calculated extrinsic information is fed to the channel decoder. After channel decoding, the decoder outputs extrinsic LLR to equalizer and next iteration starts.

3.4.3 Complexity Reduction

For MF approximation, when calculating the marginals of different frequency-domain symbols, the messages are computed once per block. Therefore, the detection complexity for the proposed MF approximation method is $\mathcal{O}(N^2)$ for a N-length symbol vector. For Bethe approximation, since $\tilde{m}_{\backslash i, x_{f,j}}$ and $\tilde{v}_{\backslash i, x_{f,j}}$ have to be calculated for different symbols, an additional complexity of $\mathcal{O}(N^2)$ is required. Consequently, the detection complexity of Bethe approximation is $\mathcal{O}(2N^2)$. In fact, $\tilde{q}_{\backslash i}(x_{f,j})$ for different symbols only differ in one term compared with its marginal

$q(x_{f,j})$. a low-complexity scheme is proposed which employs $q(x_{f,j})$ to approximate $\tilde{q}_{\backslash i}(x_{f,j})$. As a result, the complexity of the Bethe approximation reduces to $\mathcal{O}(N^2)$. this approximation method is referred to as "app-Bethe" which will be evaluated via simulations.

To further reduce the complexity in practical receivers, only the interferences from only R frequency-domain symbols are considered, whose indices correspond to the entries of $\mathbf{C}_{:,i}$ with relatively large modulus. Accordingly, the detection complexity is reduced to $\mathcal{O}(NR)$ for both MF approximation and app-Bethe method and $\mathcal{O}(2NR)$ for the standard Bethe approximation, which means the complexity increases linearly with the block length N.

3.4.4 Simulation Results

An LDPC coded FTN system with coding rate $R = 1/2$ and block length $N = 1000$ is considered . Root-raised-cosine pulse shaping with roll-off factor $\alpha = 0.5$ is employed. The number of CP symbols is $2M = 20$. The carrier frequency is $f_0 = 2$ GHz, symbol duration is $T = 0.25\mu$s, and mobile velocity is $v = 300$ km/hr. The number of channel taps $L = 10$ and the channel coefficients vary symbol by symbol. For each channel realisation, $h_{k,l}$ are independently generated according to the distribution $\mathcal{CN}(h_{k,l}, 0, q^l)$. The power delay profile is $q^l = \exp(-0.1l)$ and the averaged channel energy is normalised to 1. The number of iterations between channel decoding and equalisation is 20.

Bit error rate (BER) performances of the equalizer with the proposed algorithms and the iterative FDE-MMSE equalizer are illustrated in Fig. 3.2. For comparison purpose, the performance of a TDE method extended from [54] is also included. The number of interfering frequency-domain symbols for both MF and Bethe approximations is set to $R = 10$. It can be observed that the performance of the proposed algorithms for FTN signaling attain that of the Nyquist signaling while increas-

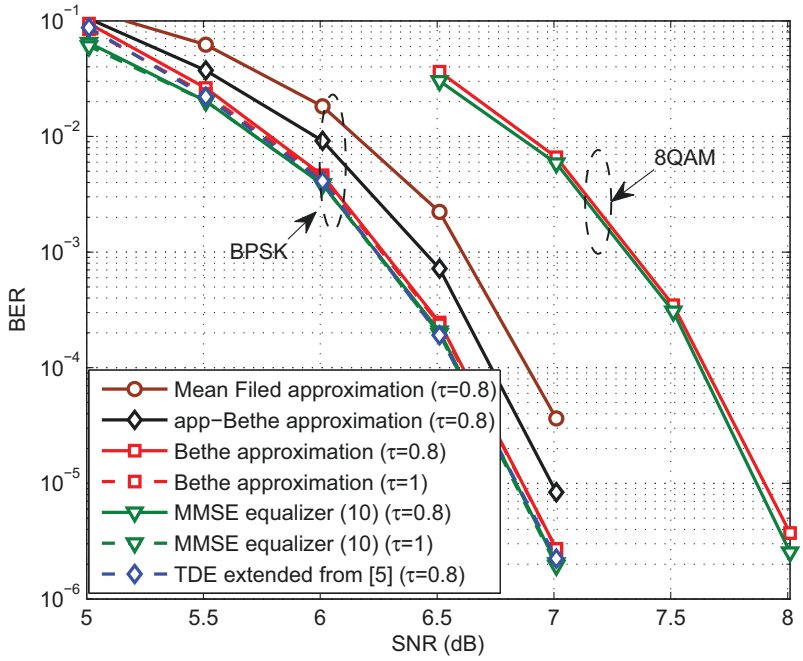


Figure 3.2 : BER performance of different algorithms for FTN signaling ($\tau = 0.8$) and Nyquist signaling ($\tau = 1$) in DSCs.

ing the transmission rate by 25%. It is seen that, with much lower computational complexity than MMSE equalizer, the Bethe approximation performs very close to the FDE-MMSE equalizer and the TDE method[1]. The MF approximation leads to performance loss due to the underestimation of variances of detected symbols. The performance of app-Bethe approximation is between that of MF and Bethe approximations, which achieves a compromise between BER performance and complexity.

BER performance of the proposed MF and Bethe approximations with different R are evaluated in Fig. 3.3. The packing ratio of FTN signaling $\tau = 0.8$. It is observed that increasing R improves the performance at the cost of higher complexity. Nevertheless, when R is greater than 10, the improvement becomes marginal.

[1]The complexity of TDE extended from [54] is $\mathcal{O}(N((L_{\mathrm{FTN}}+p)^2+L^2))$, where L_{FTN} is the length of ISI imposed by FTN and p is the order of AR model for colored noise process.

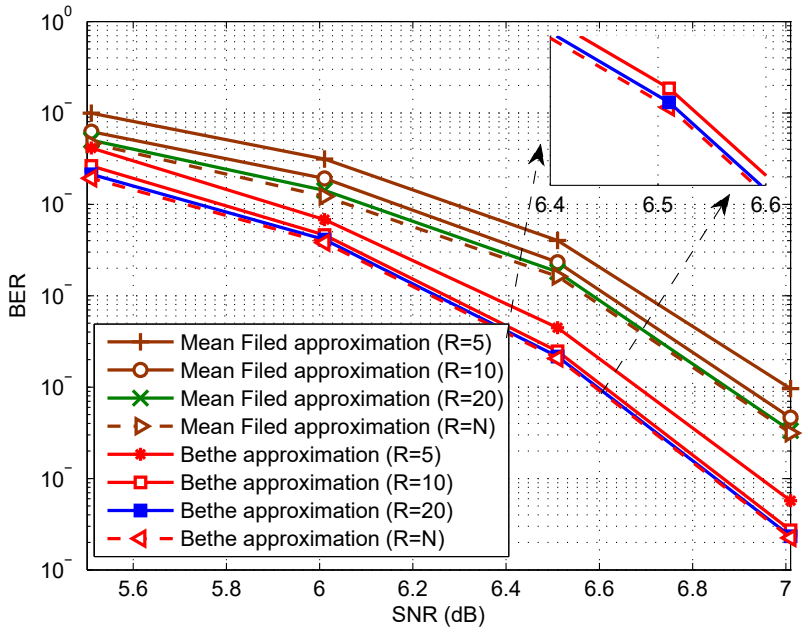


Figure 3.3 : BER performance of MF and Bethe approximations versus different values of R for FTN signaling ($\tau = 0.8$) in DSCs.

Therefore, in practical applications, good tradeoff off between BER performance and computational complexity can be achieved by choosing a proper value of R for the proposed algorithms.

3.5 Iterative Message Passing Receiver for FTN signaling in DSCs

3.5.1 Factor Graph Model

Inspired by [83], $\mathbf{C}$ is approximated by considering the interference only from the $Q-1$ neighboring frequency-domain symbols. Accordingly, the kth row of $\mathbf{C}$ becomes $\mathbf{C}_{k,:} = [0, ..., 0, \mathbf{C}_{k,k-Q+1}, ..., \mathbf{C}_{k,k}, 0, ..., 0]$ and $r_{f,k}$ is formulated as

$$r_{f,k} = \sum_{j=0}^{Q-1} \mathbf{C}_{k,k-j} \lambda_{k-j} x_{f,k-j} + \gamma_{f,k}. \tag{3.24}$$

By defining $\mathbf{s}_k \triangleq [\lambda_{k-Q+1}x_{f,k-Q+1}, ..., \lambda_k x_{f,k}]^T$ and $\mathbf{c}_k \triangleq [\mathbf{C}_{k,k-Q-1}, ..., \mathbf{C}_{k,k}]^T$, the linear state-space model is given as

$$\mathbf{s}_k = \mathbf{A}\mathbf{s}_{k-1} + \mathbf{b}_k x_{f,k}, \tag{3.25}$$

$$r_{f,k} = \mathbf{c}_k^T \mathbf{s}_k + \gamma_{f,k}, \tag{3.26}$$

where $\mathbf{A} = \begin{bmatrix} \mathbf{0}_{Q-1}^T & \mathbf{I}_{Q-1} \\ & \mathbf{0}_Q \end{bmatrix}$, $\mathbf{b}_k = [\mathbf{0}_{Q-1}, \lambda_k]^T$. Based on (3.25) and (3.26), the Forney-style factor graph (FFG) for FTN signaling in DSC is illustrated in Fig. 3.4. With Gaussian assumption of the data symbols, all messages on FFG can be expressed in Gaussian form with means and covariance matrices updated by Gaussian message passing (GMP). The notation $\overrightarrow{\cdot}$ is used to denote the message passed along the arrow direction on FFG while $\overleftarrow{\cdot}$ the message passed in opposite direction.

3.5.2 FDE-based Message Passing Receiver

The decoder relies on the parameters update of two messages on the factor graph, i.e., the mean and covariance matrix of incoming message $\overrightarrow{\mathbf{m}}_{\mathbf{x}_f} = [\overrightarrow{m}_{x_{f,0}}, ..., \overrightarrow{m}_{x_{f,N-1}}]^T$, $\overrightarrow{\mathbf{V}}_{\mathbf{x}_f} = \text{diag}\{\overrightarrow{V}_{x_{f,0}}, ..., \overrightarrow{V}_{x_{f,N-1}}\}$ and that of the outgoing message $\overleftarrow{\mathbf{m}}_{\mathbf{x}_f} = [\overleftarrow{m}_{x_{f,0}}, ..., \overleftarrow{m}_{x_{f,N-1}}]^T$, $\overleftarrow{\mathbf{V}}_{\mathbf{x}_f} = \text{diag}\{\overleftarrow{V}_{x_{f,0}}, ..., \overleftarrow{V}_{x_{f,N-1}}\}$.

- *Calculation of the incoming messages:*

Following the update rules of GMP, the parameters of incoming message can be calculated by

$$\overrightarrow{\mathbf{m}}_{\mathbf{x}_f} = \mathbf{F}\overrightarrow{\mathbf{m}}_{\mathbf{x}}, \tag{3.27}$$

$$\overrightarrow{\mathbf{V}}_{\mathbf{x}_f} = \mathbf{F}\overrightarrow{\mathbf{V}}_{\mathbf{x}}\mathbf{F}^H, \tag{3.28}$$

where $\overrightarrow{\mathbf{m}}_{\mathbf{x}}$ and $\overrightarrow{\mathbf{V}}_{\mathbf{x}}$ are the mean and covariance matrix of time-domain symbols, which can be obtained by using the extrinsic log-likelihood ratio (LLR) obtained

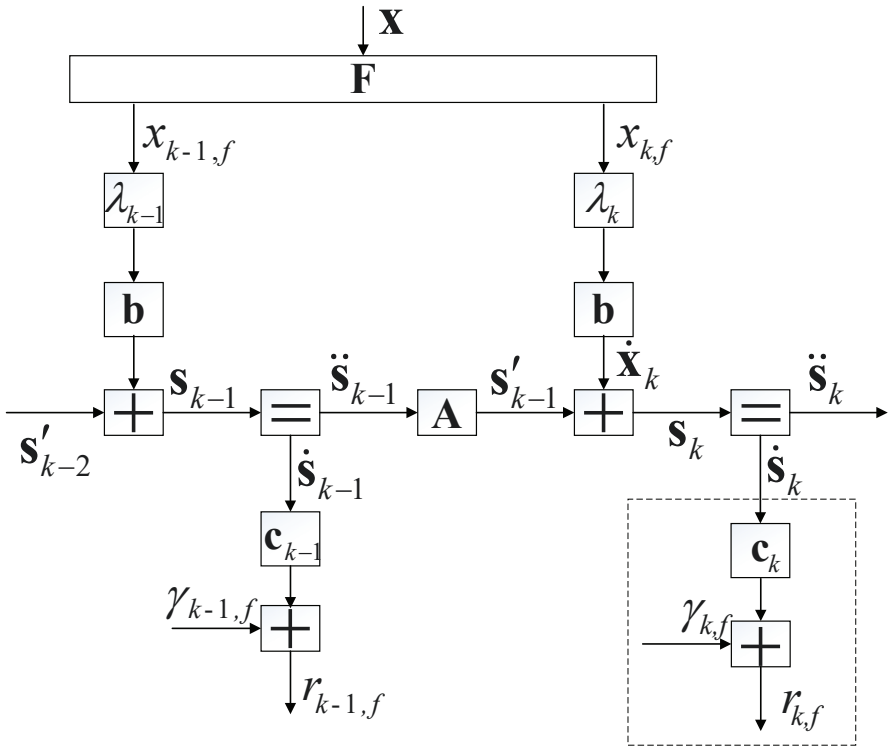


Figure 3.4 : Factor graph of FDE-based receiver for FTN signaling in DSC.

from the output of channel decoder. The calculation in (3.28) is not trivial. To reduce the complexity, the approximation in [93] is employed

$$\overrightarrow{\mathbf{V}}_{\mathbf{x}} \approx a\mathbf{I}, \quad \text{with } a = \text{tr}[\overrightarrow{\mathbf{V}}_{\mathbf{x}}]/N. \tag{3.29}$$

Accordingly, $\overrightarrow{\mathbf{V}}_{\mathbf{x}_f} = a\mathbf{I}$.

- *Calculation of the outgoing messages:*

Assuming the parameters $\overrightarrow{\mathbf{m}}_{\mathbf{s}'_{k-2}}$ and $\overrightarrow{\mathbf{V}}_{\mathbf{s}'_{k-2}}$ are available, we have

$$\overrightarrow{\mathbf{V}}_{\mathbf{s}_{k-1}} = \overrightarrow{\mathbf{V}}_{\mathbf{s}'_{k-2}} + \mathbf{b}|\lambda_{k-1}|^2 \overrightarrow{V}_{x_{f,k-1}} \mathbf{b}^T, \tag{3.30}$$

$$\overrightarrow{\mathbf{m}}_{\mathbf{s}_{k-1}} = \overrightarrow{\mathbf{m}}_{\mathbf{s}'_{k-2}} + \mathbf{b}\lambda_{k-1} \overrightarrow{m}_{x_{f,k-1}}. \tag{3.31}$$

Then the transformed mean and covariance matrix $\overrightarrow{\mathbf{W}}_{\mathbf{s}'_{k-1}} \overrightarrow{\mathbf{m}}_{\mathbf{s}'_{k-1}}$ and $\overrightarrow{\mathbf{V}}_{\mathbf{s}'_{k-1}}$ can be obtained as

$$\overrightarrow{\mathbf{V}}_{\mathbf{s}'_{k-1}} = \mathbf{A}\left(\overleftarrow{\mathbf{W}}_{\dot{\mathbf{s}}_{k-1}} + \overrightarrow{\mathbf{W}}_{\mathbf{s}_{k-1}}\right)^{-1}\mathbf{A}^T, \tag{3.32}$$

$$\overrightarrow{\mathbf{W}}_{\mathbf{s}'_{k-1}}\overrightarrow{\mathbf{m}}_{\mathbf{s}'_{k-1}} = \mathbf{A}\left(\overrightarrow{\mathbf{W}}_{\mathbf{s}_{k-1}}\overrightarrow{\mathbf{m}}_{\mathbf{s}_{k-1}} + \overleftarrow{\mathbf{W}}_{\dot{\mathbf{s}}_{k-1}}\overleftarrow{\mathbf{m}}_{\dot{\mathbf{s}}_{k-1}}\right), \tag{3.33}$$

where $\overrightarrow{\mathbf{W}}_{\mathbf{s}_{k-1}} \triangleq \overrightarrow{\mathbf{V}}^{-1}_{\mathbf{s}_{k-1}}$ and

$$\overleftarrow{\mathbf{W}}_{\dot{\mathbf{s}}_{k-1}} = \frac{\mathbf{c}_{k-1}\mathbf{c}^H_{k-1}}{N_0\lambda_{k-1}}, \tag{3.34}$$

$$\overleftarrow{\mathbf{W}}_{\dot{\mathbf{s}}_{k-1}}\overleftarrow{\mathbf{m}}_{\dot{\mathbf{s}}_{k-1}} = \frac{\mathbf{c}_{k-1}r_{f,k-1}}{N_0\lambda_{k-1}}. \tag{3.35}$$

Similarly, the backward mean and covariance matrix $\overleftarrow{\mathbf{m}}_{\mathbf{s}_k}$ and $\overleftarrow{\mathbf{V}}_{\mathbf{s}_k}$ can be expressed as

$$\overleftarrow{\mathbf{V}}_{\mathbf{s}_k} = \left(\overleftarrow{\mathbf{W}}_{\dot{\mathbf{s}}_k} + \overleftarrow{\mathbf{W}}_{\ddot{\mathbf{s}}_k}\right)^{-1}, \tag{3.36}$$

$$\overleftarrow{\mathbf{m}}_{\mathbf{s}_k} = \overleftarrow{\mathbf{V}}_{\mathbf{s}_k}\big(\overleftarrow{\mathbf{W}}_{\dot{\mathbf{s}}_k}\overleftarrow{\mathbf{m}}_{\dot{\mathbf{s}}_k} + \overleftarrow{\mathbf{W}}_{\ddot{\mathbf{s}}_k}\overleftarrow{\mathbf{m}}_{\ddot{\mathbf{s}}_k}\big). \tag{3.37}$$

Based on (3.32), (3.33), (3.36) and (3.37), we have

$$\overleftarrow{m}_{x_{f,k}} = \frac{\mathbf{b}_k^T\left(\overleftarrow{\mathbf{m}}_{\mathbf{s}_k} - \overrightarrow{\mathbf{m}}_{\mathbf{s}'_{k-1}}\right)}{\lambda_k}, \tag{3.38}$$

$$\overleftarrow{V}_{x_{f,k}} = \frac{\mathbf{b}_k^T\big(\overrightarrow{\mathbf{V}}_{\mathbf{s}'_{k-1}} + \overleftarrow{\mathbf{V}}_{\mathbf{s}_k}\big)\mathbf{b}_k}{\lambda_k^2}. \tag{3.39}$$

Then, the parameters of outgoing message $\overleftarrow{\mathbf{m}}_{\mathbf{x}_f}$, $\overleftarrow{\mathbf{V}}_{\mathbf{x}_f}$ can be obtained using (3.38) and (3.39).

By inverse DFT, the extrinsic information of equalizer can be obtained, which are used to calculate the LLRs fed to the channel decoder. After channel decoding, the extrinsic LLRs of time-domain symbols are passed to equalizer, which begins the next iteration.

3.5.3 Imperfect Frequency-Domain Channel Information

In practice, the channel information $\mathbf{C}$ may not be perfectly known, in which case the true channel information $\bar{\mathbf{C}}$ is given as

$$\bar{\mathbf{C}} = \rho\mathbf{C} + \mathbf{\Delta}, \tag{3.40}$$

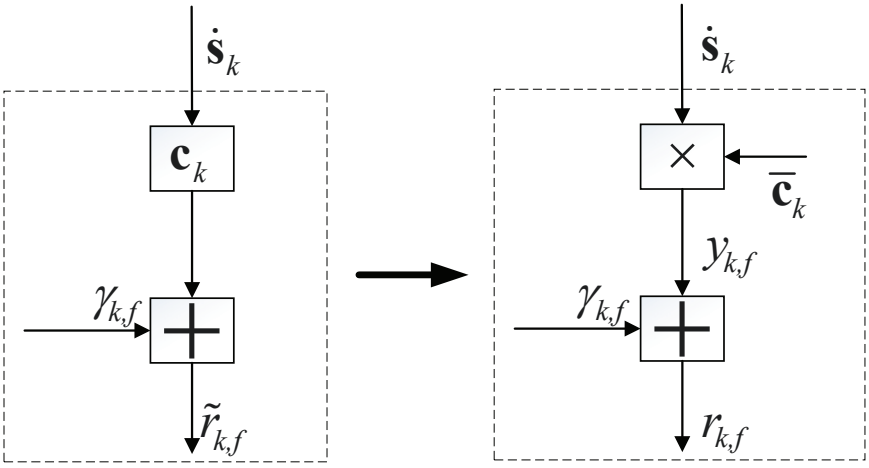


Figure 3.5 : Factor graph modification for imperfect channel information.

where ρ is the coefficient and $\mathbf{\Delta}$ is the channel uncertainty. The elements of $\mathbf{\Delta}$ are assumed to be independent Gaussian distributed variables with zero means and variances $\{\phi_{i,j}\}$. Ignoring the channel uncertainty in the above equalisation may lead to crucial performance loss. To tackle this problem, the channel uncertainty is taken into account on factor graph and robust message passing receiver for FTN signaling in DSCs is derived.

Substituting (3.40) into (3.26) yields $r_{f,k} = (\rho\mathbf{c}_k + \mathbf{\Delta}_k)^T\mathbf{s}_k + \gamma_{f,k}$, where $\mathbf{\Delta}_k \triangleq [\mathbf{\Delta}_{k,k-Q-1}, ..., \mathbf{\Delta}_{k,k}]^T$. Accordingly, the dashed-line box in Fig. 3.4 can be replaced by the one illustrated in Fig. 3.5. Note that a multiplier node $\boxtimes$ is included to represent the product of two variables. Denote $\bar{\mathbf{c}}_k = [\bar{\mathbf{C}}_{k,k-Q+1}, ..., \bar{\mathbf{C}}_{k,k}]^T$ and $\mathbf{\Phi}_k = \text{diag}\{\phi_{k,k-Q+1}, \ldots, \phi_{k,k}\}$. According to sum-product algorithm, the message from multiplier node to $\dot{\mathbf{s}}_k$ can be calculated as[2]

$$\begin{aligned}
\overleftarrow{\boldsymbol{\mu}}_{\dot{\mathbf{s}}_k} &= \int \delta(y_{f,k} - \bar{\mathbf{c}}_k^T\dot{\mathbf{s}}_k)\overleftarrow{\mu}_{y_{f,k}}\overleftarrow{\boldsymbol{\mu}}_{\bar{\mathbf{c}}_k}\mathrm{d}\, y_{f,k}\,\mathrm{d}\,\bar{\mathbf{c}}_k \\
&\propto \int e^{-\frac{1}{N_0\lambda_k}(r_{f,k}-\bar{\mathbf{c}}_k^T\dot{\mathbf{s}}_k)^2} e^{-(\bar{\mathbf{c}}_k-\rho\mathbf{c}_k)^H\mathbf{\Phi}_k^{-1}(\bar{\mathbf{c}}_k-\rho\mathbf{c}_k)}\mathrm{d}\bar{\mathbf{c}}_k, \\
&\propto \exp\left(\dot{\mathbf{s}}_k^H\frac{\rho^2\mathbf{c}_k\mathbf{c}_k^H}{N_0\lambda_k+\Sigma_s}\dot{\mathbf{s}}_k - 2\dot{\mathbf{s}}_k^H\frac{\rho\mathbf{c}_k r_{f,k}}{N_0\lambda_k+\Sigma_s}\right),
\end{aligned} \tag{3.41}$$

[2]The multiplier node $\boxtimes$ can be regarded as a factor node $\delta(y_{f,k} - \bar{\mathbf{c}}_k^T\dot{\mathbf{s}}_k)$.

where $\Sigma_s = \dot{\mathbf{s}}_k^H \mathbf{\Phi}_k \dot{\mathbf{s}}_k$. For PSK modulation,[3] $|x_{f,k}|^2 = |x_k|^2 = 1$ satisfies. Then, after tedious but straightforward manipulations, the message $\overleftarrow{\boldsymbol{\mu}}_{\dot{\mathbf{s}}_k}$ in (3.41) can be expressed with parameters given by

$$\overleftarrow{\mathbf{W}}_{\dot{\mathbf{s}}_k} \overleftarrow{\mathbf{m}}_{\dot{\mathbf{s}}_k} = \frac{\rho \mathbf{c}_k r_{f,k}}{N_0 \lambda_k + \mathrm{tr}[\mathbf{\Phi}_k]}, \tag{3.42}$$

$$\overleftarrow{\mathbf{W}}_{\dot{\mathbf{s}}_k} = \frac{\rho^2 \mathbf{c}_k \mathbf{c}_k^H}{N_0 \lambda_k + \mathrm{tr}[\mathbf{\Phi}_k]}. \tag{3.43}$$

The parameters of the other messages on factor graph can be derived similarly to that in Section 3.5.2, which are not given here for brevity.

The complexity of the proposed message passing receiver is compared to state-of-the-art algorithms in Table 3.1. Since Q can be much smaller than N, the proposed algorithm reduces the complexity of MMSE equalizer significantly. Moreover, compared to the variational inference method, the complexity incensement of the proposed algorithm is acceptable.

Table 3.1 : Complexity Analysis

Algorithm	Complexity	Value
MMSE equalizer	$\mathcal{O}(N^3)$	N is large
Proposed algorithm	$\mathcal{O}(NQ^2)$	Q is small

3.5.4 Simulation Results

In the simulation, a rate-1/2 convolutional code and BPSK modulation are employed. The number of transmitted symbols is $N = 1000$. A FTN signaling with root raised cosine pulse shaping is considered, having a roll-off factor $\alpha = 0.5$. The number of CP symbols is $2M = 40$.[4] The carrier frequency is $f_0 = 2$ GHz, symbol

[3]For QAM signal, (3.41) is not in Gaussian form and GMP cannot be applied directly.

[4]The number of CP symbols should be larger than the ISI tap length caused by both FTN signaling and fading channel.

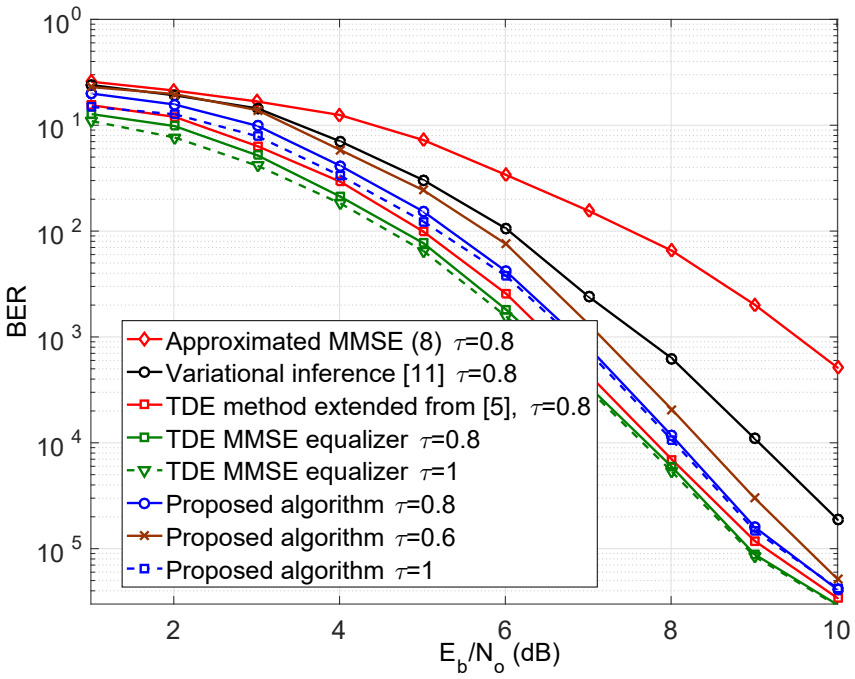


Figure 3.6 : BER performance of different algorithms for Nyquist signaling and FTN signaling.

duration is $T = 0.25\mu$s, and mobile velocity $v = 300$ km/hr. The channel varies symbol by symbol. The number of channel taps is set to $L = 30$. For each channel realisation, the coefficients $h_{k,l}$ are independently generated according to the distribution $h_{k,l} \sim \mathcal{N}(0, q^l)$. The power delay profile is $q^l = \exp(-0.1l)$ and the averaged channel energy is normalised to 1. For the proposed algorithm, $Q = 5$ is set, unless otherwise specified. The number of iterations is $I = 20$. The parameters of message from channel decoder are initialised as $\overrightarrow{\mathbf{m}}_{\mathbf{x}}^{(0)} = \mathbf{0}^T$ and $\overrightarrow{\mathbf{V}}_{\mathbf{x}}^{(0)} = \mathbf{I}_N$.

The bit error rate (BER) performance of the proposed algorithm is illustrated in Fig. 3.6, in which the channel information is assumed to be perfectly known. For comparison, the performance of MMSE equalizer[5], TDE method extended from

[5]To exploit the prior information in (3.29) obtained from channel decoder, MMSE and the approximated MMSE equalizer are modified to $\mathbf{W} = a\mathbf{\Lambda}^H\mathbf{C}^H(a\mathbf{C}\mathbf{\Lambda}\mathbf{\Lambda}^H\mathbf{C}^H + N_0\mathbf{\Lambda})^{-1}$ and $\mathbf{W}_{i,i} = a\mathbf{C}_{i,i}\lambda_i/(|a\mathbf{C}_{i,i}\lambda_i|^2 + N_0\lambda_i)$, respectively.

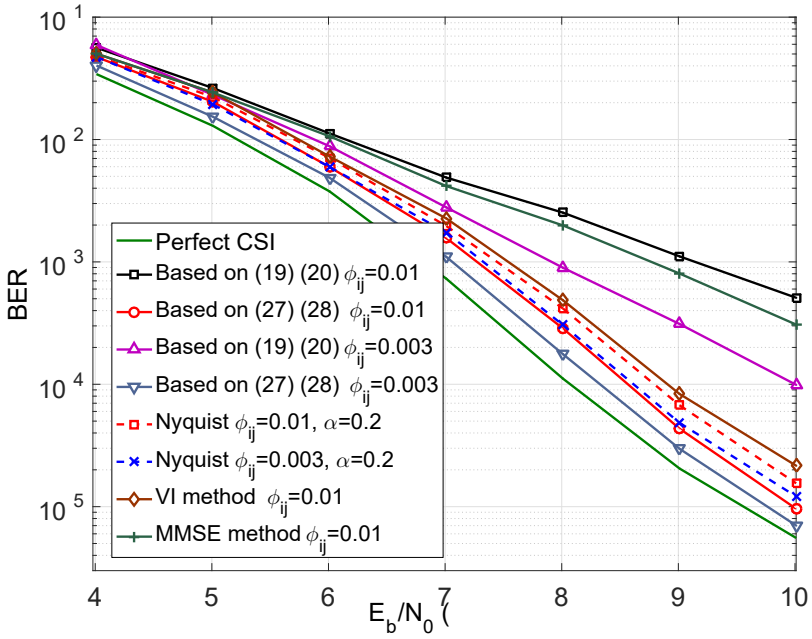


Figure 3.7 : BER performance of the proposed robust algorithm with channel information uncertainty. The parameter $\rho = 0.99$.

[54] and the variational inference method in [95] are also plotted. It is observed that the approximated MMSE equalizer suffers significant performance loss. The proposed algorithm outperforms the variational inference method by considering the conditional dependence between frequency-domain symbols. It is also observed that FTN signaling with $\tau = 0.6$ can transmit 66% above the Nyquist rate, with only 0.4dB performance loss. The proposed algorithm with imperfect channel information is evaluated in Fig. 3.7. The variance of the elements of $\mathbf{\Delta}$ are set to $\phi_{ij} = 0.003$ and $\phi_{ij} = 0.01$, respectively,[6] $\rho = 0.99$. The performance of Nyquist signaling with $\alpha = 0.2$ is plotted as a benchmark. It is seen that FTN signaling outperforms the Nyquist counterpart given the same spectral efficiency. Significant performance degradation can be observed at high E_b/N_0 if the uncertainty of channel information

[6]The variances are evaluated coarsely based on channel estimation with 10 and 30 pilot symbols, respectively, in time-invariant channel at SNR=10 dB.

is not taken into account, while the proposed robust algorithm performs very close to the one with perfect channel information. Moreover, compared to the MMSE and variational inference method, the proposed message passing receiver is more robust to channel uncertainty.

3.6 Conclusions

This Chapter proposes two FDE based iterative receivers for FTN signaling in DSCs. By minimising the constrained VFE corresponding to MF and Bethe approximations, extrinsic information from the frequency-domain equalizer is derived in a Gaussian form. By using Gaussian message passing, iterative symbol detection is performed on a reformulated vector-form factor graph. To further reduce the complexity, the number of interfering frequency-domain symbols is restricted. Accordingly, the complexity of the proposed algorithms increases linearly with the block length N. Simulation results show that the proposed algorithms for FTN signaling in DSCs attain the performance of Nyquist signaling while increasing the transmission rate by 25%. By considering the conditional dependencies of pairwise symbols, the Bethe approximation outperforms the MF method, and it performs very close to the FDE-MMSE equalizer and the extended TDE method in DSCs with much lower computational complexity. Moreover, by taking into account the channel uncertainty, the proposed message passing receiver is shown to be robust to the imperfect channel information.

Chapter 4

Uplink SCMA Multiuser Detector Design and Convergence Analysis

4.1 Introduction

In Chapter 2 and Chapter 3, the receiver designs for FTN signaling are investigated and it is shown that the spectral efficiency can be improved by using non-orthogonal waveform. In this chapter, a low complexity receiver for uplink SCMA system will be developed.

With the aid of appropriate sparse codebook design, SCMA achieves an improved performance. However, due to the non-orthogonal resource allocation of the SCMA system, the optimal maximum *a posteriori* (MAP) detectors impose a high complexity. By exploiting the sparsity of the codewords, several factor graph and message passing algorithm (MPA) [58] based receivers have been developed [59, 62, 96, 97]. Nevertheless, the rank-deficient SCMA system results in a factor graph having short cycles, for which the message passing algorithm may not be able to converge. Therefore, it is important to investigate the convergence of iterative SCMA receivers.

Hence, a low-complexity receiver based on Bayesian inference [98] is proposed in this chapter. The system model of uplink SCMA system is first presented. Then instead of conventional probabilistic factorisation, the joint posterior distribution of data symbols is formulated as the product of several clique potentials, which can help to reduce the number of short cycles. By minimising the Bethe approximation [99] based VFE, the closed forms of marginal distributions of symbols are determined. It can be seen that the complexity of the proposed receiver only increases linearly

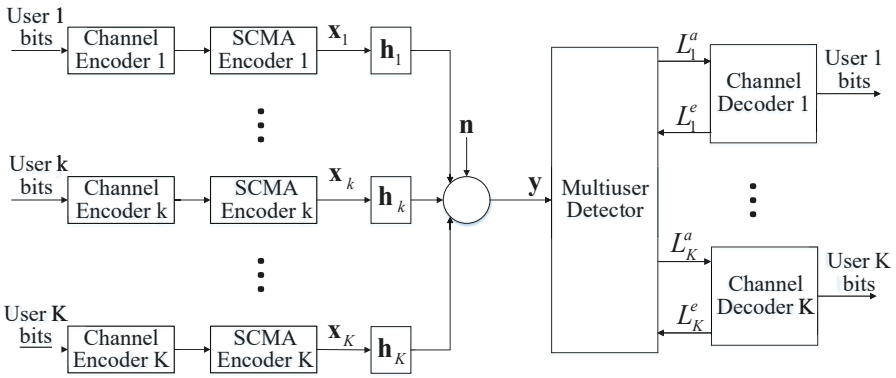


Figure 4.1 : Block diagram of the SCMA system.

with the number of users. Moreover, since the proposed scheme is an iterative one, the convergence behavior of the proposed algorithm is analysed. It is proved that the variance of symbol belief can always converge, while for the mean, the necessary and sufficient conditions to guarantee its convergence are given.

4.2 System Model

An SCMA system is considered which supports K users multiplexed onto J orthogonal resources is considered, as shown in Fig. 4.1. In orthogonal multiple access (OMA), $K \leq J$ is satisfied to ensure that each user is allocated to an orthogonal resource. By contrast, in the SCMA system, the number of users is larger than the number of resource blocks, which leads to a rank-deficient system. The ratio $\lambda = \frac{K}{J} > 1$ is referred to as the normalised user-load. For each user, the bits $\mathbf{b}_k$ are mapped to a J-component SCMA codeword $\mathbf{x}_k$. The mapping function for the kth user is $\mathbf{x}_k = \phi(\mathbf{b}_k)$, $\phi : \mathbb{B}^{\log_2 M} \rightarrow \boldsymbol{\chi}$, where $\boldsymbol{\chi} \in \mathbb{C}^K$ and $|\boldsymbol{\chi}| = M$. The SCMA codeword of the kth user is selected from a predefined codebook. Let $\mathbf{x}_k = [x_{k,1}, ..., x_{k,J}]$ be the transmitted symbols of user k. Given the sparsity of SCMA codewords, there are only $D < J$ nonzero entries in $\mathbf{x}_k$.

The spread signal of user k is then transmitted through the corresponding chan-

nel $\mathbf{h}_k = [h_{k,1}, ..., h_{k,J}]^T$. Assuming perfect synchronisation between the base station and users, the received signal is expressed as

$$\mathbf{y} = \sum_{k=1}^{K} \text{diag}(\mathbf{h}_k)\mathbf{x}_k + \mathbf{n}, \tag{4.1}$$

where $\mathbf{n}$ is the Gaussian noise having a power spectral density of N_0.

4.3 Proposed Low-Complexity Receiver

Generally, utilising message passing algorithm (MPA) based receiver can provide the optimal maximum *a posteriori* (MAP) estimate. Nevertheless, MPA has exponential complexity which limits its implementation in practice. In this section, a low-complexity receiver for SCMA system is proposed based on the variational free energy framework.

4.3.1 The Proposed Algorithm

Under the Gaussian assumption of $\mathbf{n}$, the likelihood function of the received signals $\mathbf{y}$ conditioned on the user's transmitted symbols $\mathbf{x}_k$ can be derived, formulating as

$$p(\mathbf{y}|\mathbf{x}) \propto \exp\left(-\frac{1}{N_0}\|\mathbf{y} - \sum_{k=1}^{K} \text{diag}(\mathbf{h}_k)\mathbf{x}_k\|^2\right). \tag{4.2}$$

According to the classic Bayesian theorem, the *a posteriori* distribution can be expressed as

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x})p(\mathbf{y}|\mathbf{x}), \tag{4.3}$$

where $p(\mathbf{x})$ is the joint *a priori* distribution of all users' source symbols, which is expressed as

$$p(\mathbf{x}) \propto \prod_{k=1}^{K} \mathcal{N}(\mathbf{x}_k; \mathbf{m}^0_{x_k}, \mathbf{V}^0_{x,k}). \tag{4.4}$$

In (4.4), $\mathbf{m}_{x_k}^0 = [m_{x_{k,1}}^0, ..., m_{x_{k,J}}^0]^T$ and $\mathbf{V}_{x,k}^0 = \text{diag}([v_{x_{k,1}}^0, ..., v_{x_{k,J}}^0])$ are calculated based on the extrinsic information gleaned from the channel decoder.

Since the SCMA source symbols and received signal samples are independent, (4.3) is factorised as

$$p(\mathbf{x}|\mathbf{y}) \propto \prod_{j=1}^{J} \prod_{k} \phi_k^j \prod_{(i,k)} \phi_{i,k}^j, \tag{4.5}$$

where

$$\phi_k^j = p(x_{k,j}) \exp\left(-\frac{h_{k,j}^2 x_{k,j}^2 - 2h_{k,j} y_j x_{k,j}}{N_0}\right), \tag{4.6}$$

$$\phi_{i,k}^j = \phi_{k,i}^j = \exp\left(-\frac{2h_{k,j}^2 x_{i,j} x_{k,j}}{N_0}\right), \tag{4.7}$$

are referred to as the self-potential and pairwise potential derived in [100], respectively. Usually, the goal is to find the estimate of every data symbol, i.e. $\hat{x}_{k,j}$, which is equivalent to obtain the *a posteriori* probability $p(x_{k,j}|\mathbf{y})$. Direct marginalisation of $p(\mathbf{x}|\mathbf{y})$ imposes a high complexity. Motivated by the energy minimisation framework, an appropriate trail distribution $b(\mathbf{x})$ is needed that can be readily marginalised to approximate $p(\mathbf{x}|\mathbf{y})$.

The variational free energy is defined as [101]

$$F = F_H + \int b(\mathbf{x}) \ln \frac{b(\mathbf{x})}{p(\mathbf{x}|\mathbf{y})} \mathrm{d}\mathbf{x}, \tag{4.8}$$

where $F_H = -\ln Z$ is termed as *Helmholtz free energy* [101]. In this problem, Z denotes the normalisation factor of $p(\mathbf{x}|\mathbf{y})$. The simplest form for $b(\mathbf{x})$ is the mean-field (MF) approach [100], which factorises $b(\mathbf{x})$ as $b_{\text{MF}}(\mathbf{x}) = \prod_{k,j} b_k^j(x_{k,j})$, where $b_k^j(x_{k,j})$ is a marginalised trail distribution ('belief') over the single variable $x_{k,j}$. Accordingly, the free energy can be readily computed and then an MF approximation for the beliefs $b_k^j(x_{k,j})$ is obtained. However, the main problem of the factorised MF approximation is that it assumes all variables in $\mathbf{x}$ to be conditionally independent of each other, even though actually they are not. This motivates us to find a

more accurate approximation than the MF approximation. The Bethe method has been recognised as an efficient tool in probabilistic problems, since it considers the conditional dependencies amongst the variables as follows:

$$b(\mathbf{x}) = \prod_j \prod_k b_k^j(x_{k,j}) \prod_{(i,k)} \frac{b_{i,k}^j(x_{k,j}, x_{i,j})}{b_k^j(x_{k,j}) b_i^j(x_{i,j})}. \tag{4.9}$$

Substituting (4.9) into (4.8) yields

$$\begin{aligned} F = & \sum_j \left(\sum_{(i,k)} \int b_{i,k}^j(x_{k,j}, x_{i,j}) \ln \frac{b_{i,k}^j(x_{k,j}, x_{i,j})}{\phi_{i,k}^j} \mathrm{d}x_{k,j} \mathrm{d}x_{i,j} \right. \\ & \left. + (K-1) \sum_k \int b_k^j(x_{k,j}) \ln \frac{b_k^j(x_{k,j})}{\phi_k^j} \mathrm{d}x_{k,j} \right) + C. \end{aligned} \tag{4.10}$$

The variational free energy is constrained by the normalisation and consistency constraints of:

$$\int b_k^j(x_{k,j}) \mathrm{d}x_{k,j} = 1, \tag{4.11}$$

$$\int b_{i,k}^j(x_{k,j}, x_{i,j}) \mathrm{d}x_{i,j} = b_k^j(x_{k,j}). \tag{4.12}$$

The classic Lagrangian multipliers λ_k and $\lambda_{k,i}(x_{k,j})$ are used for the normalisation constraint and the consistency constraints. Thus the Lagrangian is constructed as

$$\begin{aligned} L = & F + \sum_j \left(\sum_i \lambda_k \left(1 - \int b_k^j(x_{k,j}) \mathrm{d}x_{k,j} \right) + \right. \\ & \left. \sum_{(i,k)} \int \lambda_{i,k}(x_{k,j}) \left(\int b_{i,k}^k(x_{k,j}, x_{i,j}) \mathrm{d}x_{i,j} - b_k^j(x_{k,j}) \right) \mathrm{d}x_{k,j} \right). \end{aligned} \tag{4.13}$$

According to the calculus of variations, setting $\delta(L) = 0$ gives the beliefs $b_{i,k}^j(x_{k,j}, x_{i,j})$ and $b_k^j(x_{k,j})$ at stationary points, which yield

$$b_{i,k}^k(x_{k,j}, x_{i,j}) = C \phi_{i,k}^j b_i^j(x_{i,j}) b_k^j(x_{k,j}) \exp\left[-\lambda_{k,i}(x_{k,j}) - \lambda_{i,k}(x_{i,j}) \right], \tag{4.14}$$

$$b_k^j(x_{k,j}) = C \phi_k^j \exp\left(\sum_{(i,k)} \lambda_{i,k}(x_{k,j}) \right). \tag{4.15}$$

By substituting (4.15) into (4.14), $b_{i,k}^j(x_{k,j}, x_{i,j})$ can be expressed as

$$b_{i,k}^j(x_{k,j}, x_{i,j}) \propto \phi_k^j \phi_i^j \phi_{i,k}^j \exp\left(\sum_{(k,l), l \neq i} \lambda_{k,l}(x_{k,j}) + \sum_{(i,m), m \neq k} \lambda_{i,k}(x_{i,j}) \right).$$

For simplicity, the following shorthand is introduced

$$b^j_{i\backslash k}(x_{i,j}) \propto \phi^j_i \exp\left(\sum_{(i,m),m\neq k} \lambda_{i,m}(x_{i,j})\right). \tag{4.16}$$

After integrating $b^j_{i,k}(x_{k,j}, x_{i,j})$ over $x_{i,j}$ and comparing it to (4.15), finally $\lambda_{k,i}(x_{k,j})$ is given as

$$\lambda_{k,i}(x_{k,j}) = \ln\left(\int \phi^j_{i,k} b^j_{i\backslash k}(x_{i,j})\mathrm{d}x_{i,j}\right). \tag{4.17}$$

Without loss of generality, the mean and variance of $b^j_{i\backslash k}(x_{i,j})$ are denoted by $m^j_{i\backslash k}$ and $v^j_{i\backslash k}$. Then $\lambda_{k,i}(x_{k,j})$ is expressed as a quadratic polynomial

$$\begin{aligned}\lambda_{k,i}(x_{k,j}) &= \frac{h^4_{k,j} v^j_{i\backslash k}}{N_0^2} x^2_{k,j} - 2\frac{h^2_{k,j} m^j_{i\backslash k}}{N_0} x_{k,j} + C \\ &= -\beta^j_{k,i} x^2_{k,j} + 2\gamma^j_{k,i} x_{k,j} + C. \end{aligned} \tag{4.18}$$

Based on (4.15) and (4.18), the mean and variance of the approximate marginal $b^j_k(x_{k,j})$ can be calculated as

$$m_{x_{k,j}} = v_{x_{k,j}}\left(\frac{m^0_{x_{k,j}}}{v^0_{x_{k,j}}} + \frac{2h_{k,j}y_j}{N_0} + \sum_{(i,k)} \gamma^j_{k,i}\right), \tag{4.19}$$

$$v_{x_{k,j}} = \left(\frac{1}{v^0_{x_{k,j}}} + \frac{h^2_{k,j}}{N_0} + \sum_{(i,k)} \beta^k_{k,i}\right)^{-1}. \tag{4.20}$$

Since the terms $\frac{1}{v^0_{x_{k,j}}} + \frac{h^2_{k,j}}{N_0}$ and $\frac{m^0_{x_{k,j}}}{v^0_{x_{k,j}}} + \frac{2h_{k,j}y_j}{N_0}$ do not change during the iterations, ρ^j_k and ϱ^j_k are used to denote those two terms for simplicity. A damping scheme is further considered, which is expected to improve the performance in a high density connected network [102]. Denoting the belief obtained at the lth iteration as $b^j_k(l)$, the damped belief is computed as

$$\tilde{b}^j_k(l) = (b^j_k(l))^\alpha (b^j_k(l-1))^{(1-\alpha)}, \tag{4.21}$$

where $0 < \alpha < 1$ is the damping factor. That is to say the updating of the belief is based on a combination of the new belief and the old one. The mean and variance

of the damped belief are given as

$$\tilde{m}_{x_{k,j}}(l) = \tilde{v}_{x_{k,j}}(l)\left(\frac{\alpha m_{x_{k,j}}(l)}{v_{x_{k,j}}(l)} + \frac{(1-\alpha)m_{x_{k,j}}(l-1)}{v_{x_{k,j}}(l-1)}\right), \tag{4.22}$$

$$\tilde{v}_{x_{k,j}}(l) = \left(\frac{\alpha}{v_{x_{k,j}}(l)} + \frac{1-\alpha}{v_{x_{k,j}}(l-1)}\right)^{-1}. \tag{4.23}$$

Given the mean and variance of $\tilde{b}_k^j(x_{k,j})$, the log likelihood ratios (LLRs) L_k^a fed to the channel decoder can be calculated. Note that updating $m_{x_{k,j}}$ and $v_{x_{k,j}}$ relies on other variables, hence the extrinsic information corresponding to different data symbols is updated iteratively. After decoding, the output LLRs L_k^e are converted to $p(x_{k,j})$ and the next turbo iteration starts. The details of the proposed algorithm is summarised in Algorithm 2.

Algorithm 2 Energy Minimisation Based Turbo Detector for SCMA System

1: **Initialisation**:
2: The initial *a priori* distribution of user's source symbol is set as Gaussian distribution with zero mean and infinite variance;
3: **for** iter=1:N_{Iter} (number of iterations) **do**
4: For all users $\forall k$
5: Determine $\beta_{k,i}^j$ and $\gamma_{k,i}^j$ according to (4.18);
6: Compute $b_{i\backslash k}^j(x_{i,j})$ and $b_k^j(x_{k,j})$ according to (4.16) and (4.19), (4.20);
7: Update the mean and variance of the damped belief according to (4.22), (4.23);
8: Calculate LLR L_k^a based on $\tilde{b}_k^j(x_{k,j})$ and feed them to the channel decoder;
9: Perform standard BP channel decoding;
10: Calculate $m_{x_{k,j}}^0$ and $v_{x_{k,j}}^0$ based on the output extrinsic information from the channel decoder ;
11: **end for**

4.3.2 Computational Complexity

Note that the complexity of the proposed algorithm is dominated by the integration in (4.17). For the symbol $x_{k,j}$ of user k, there are a total of $(D-1)$ interfering symbols. Note that with the beliefs being Gaussian, when obtaining the marginal of symbol $x_{k,j}$, only simple addition and multiplication calculations are involved. As a result, a complexity on the order of $\mathcal{O}[(D-1)])$ is imposed and the complexity of the proposed algorithm is then $\mathcal{O}[J(D-1)]$. The original MPA receiver requires an optimal MAP detector, having a complexity order of $\mathcal{O}(|\boldsymbol{\chi}|^D)$. The Max-log based SCMA detector of [62] still has a complexity order of $|\boldsymbol{\chi}|^D$, although the number of operations has been significantly reduced. By contrast, the complexity of the proposed algorithm only increases linearly with the number of users. Compared to the method of [97], the proposed algorithm has the same complexity order. In summary, the different algorithms' complexities summarised in Table I.

Table 4.1 : Complexity Comparison

Algorithm Name	Computational Complexity
Conventional MPA	$\mathcal{O}(K\|\boldsymbol{\chi}\|^D)$
Max-log based MPA	$\mathcal{O}(K\|\boldsymbol{\chi}\|^D)$
Modified MPA [97]	$\mathcal{O}[K\|\boldsymbol{\chi}\|(D-1)]$
The proposed algorithm	$\mathcal{O}[K\|\boldsymbol{\chi}\|(D-1)]$

4.4 Convergence Analysis

The convergence is a key issue for an iterative algorithm. In this section, the realistic conditions that guaranteed the convergence of the proposed energy minimisation based receiver are derived.

According to (4.16), at the lth iteration, $m^j_{i\backslash k}$ and $v^j_{i\backslash k}$ are updated based on the parameters determined in the previous iteration, following

$$v^j_{i\backslash k}(l) = \left(\rho^j_i + \sum_{(i,m),m\neq k} \beta^j_{i,m}(l-1) \right)^{-1}, \tag{4.24}$$

$$m^j_{i\backslash k}(l) = v^j_{i\backslash k}(l) \left(\varrho^j_i + \sum_{(i,m),m\neq k} \gamma^j_{i,m}(l-1) \right). \tag{4.25}$$

Proposition 1:

The variance $\tilde{v}_{x_{k,j}}$ of the belief is guaranteed to converge, satisfying

$$\tilde{v}_{x_{k,j}}(l) \leq \tilde{v}_{x_{k,j}}(l-1).$$

Proof:

See Appendix C. ■

Next let's analyse the convergence of $\tilde{m}_{x_{k,j}}$, which is equivalent to proving that the difference between $m_{x_{k,j}}$ obtained in two consecutive iterations becomes smaller as l becomes larger, yielding,

$$|\tilde{m}_{x_{k,j}}(l+1) - \tilde{m}_{x_{k,j}}(l)| \leq |\tilde{m}_{x_{k,j}}(l) - \tilde{m}_{x_{k,j}}(l-1)|. \tag{4.26}$$

Provided that the number of iterations is large enough, the parameters, $v_{x_{k,j}}$ and $v^j_{i\backslash k}$ are assumed to converge to v^* and $\bar{v}^*$ for all k, respectively. According to (4.23), $\tilde{v}_{x_{k,j}}$ also converges to v^*. Thus (4.19) and (4.22) can be rewritten as

$$m_{x_{k,j}}(l) = v^* \left(\varrho^j_k + \sum_{(i,k)} \gamma^j_{k,i}(l) \right), \tag{4.27}$$

$$\tilde{m}_{x_{k,j}}(l) = \alpha m_{x_{k,j}}(l) + (1-\alpha) m_{x_{k,j}}(l-1). \tag{4.28}$$

Then

$$\tilde{m}_{x_{k,j}}(l+1) - \tilde{m}_{x_{k,j}}(l) = \alpha(m_{x_{k,j}}(l+1) - m_{x_{k,j}}(l)) + (1-\alpha)(m_{x_{k,j}}(l) - m_{x_{k,j}}(l-1))$$
$$= v^* \sum_{(i,k)} \left[\alpha \left(\gamma^j_{k,i}(l+1) - \gamma^j_{k,i}(l)\right) + (1-\alpha)\left(\gamma^j_{k,i}(l) - \gamma^j_{k,i}(l-1)\right)\right], \tag{4.29}$$

which implies that the convergence of $m_{x_{k,j}}$ is related to that of $\gamma^j_{k,i}$.

Substituting $m^j_{i\backslash k}$ from (4.25) into (4.18) yields

$$\gamma^j_{i,m}(l) = -b\left(\varrho^j_i + \sum_{(i,m),m\neq k} \gamma^j_{i,m}(l-1)\right), \tag{4.30}$$

with $b = \frac{h^2_{k,j}\tilde{v}^*}{N_0}$. Similar to (C.2) in Appendix C, the following inequality is obtained

$$\gamma^j_{i,m}(l+1) - \gamma^j_{i,m}(l) = b \sum_{(i,m),m\neq k} \left[\gamma^j_{i,m}(l-1) - \gamma^j_{i,m}(l)\right]. \tag{4.31}$$

Again, by stacking all γ having the index j as a vector, the following equation is given

$$\boldsymbol{\gamma}^j(l+1) - \boldsymbol{\gamma}^j(l) = b\mathbf{A}\left[\boldsymbol{\gamma}^j(l) - \boldsymbol{\gamma}^j(l-1)\right]. \tag{4.32}$$

Proposition 2:

The mean $m_{x_{k,j}}$ of the symbol belief is guaranteed to converge, if and only if the spectral radius of $\mathbf{A}$ satisfies $\rho(\mathbf{A}) < \frac{1}{b}$.

Proof:

See Appendix D. ■

4.5 Simulation Results

A pair of SCMA systems are considered using parameters of: a) $K = 6$ and $J = 4$, $\lambda = 150\%$, where the corresponding codebook is defined in [103]; b) $K = 12$

and $J = 6$, $\lambda = 200\%$, where the corresponding codebook is defined in [104]. A rate-5/7 LDPC code is employed and a flat-fading Rayleigh channel associated with perfect channel state information (CSI) is used. The number of iterations between the channel decoder and the multiuser detector is set to $N_{\text{Iter}} = 10$.

The bit error rate (BER) performance of the proposed algorithm is compared to that of the state-of-the-art methods in Fig. 4.2 and Fig. 4.3. The damping factor of the proposed algorithm is $\alpha = 0.3$.[1] It can be seen from both figures that the proposed algorithm matches the performance of the MPA receiver, but the complexity of the conventional MPA receiver increases exponentially with the number of users. The variational inference method has a low complexity as a benefit of the mean-field approximation. Nevertheless, due to ignoring the conditional dependencies amongst data symbols, the variational inference method leads to a significant performance loss. When $\lambda = 150\%$, the modified MPA method of [97] has a similar performance to that of the proposed algorithm. However, when $\lambda = 200\%$, the modified MPA [97] experiences performance loss. This is because the factor graph has more cycles due to the more severe interference, which will result in convergence problems for the modified MPA.

To evaluate the convergence of the proposed algorithm, in Fig. 4.4, its BER performance for a single user versus the number of iterations at different values of E_b/N_0 is depicted. An SCMA system associated with $\lambda = 150\%$ is considered. It can be observed that increasing the number of iterations improves the performance of the proposed algorithm. Moreover, at $E_b/N_0 = 3$ dB, the performance improvement becomes marginal, as the number of iterations increases. When E_b/N_0 becomes higher, more iterations are required to guarantee convergence. The extrinsic information transfer (EXIT) chart is additionally invoked to reveal the mutual information (MI)

[1] There are several methods proposed to find the optimal damping factor under different criteria. Here the value in [102] is used.

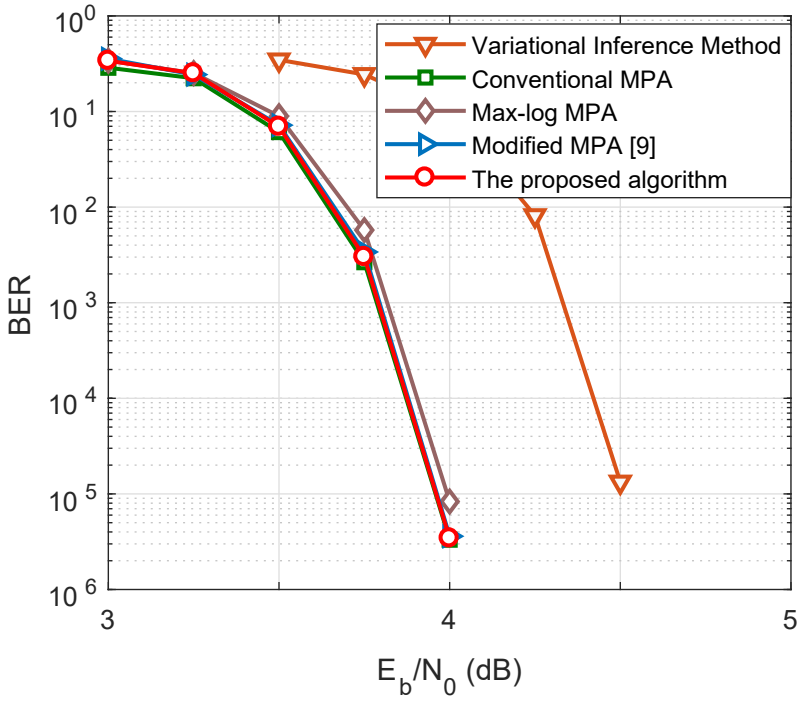


Figure 4.2 : BER performance of different algorithms. ($\lambda = 150\%$)

convergence property between the channel decoder and the proposed SCMA detector, which is shown in Fig. 4.5. $I_{A,dec}$ and to $I_{E,dec}$ are used to denote the MI between the transmitted bits and the LLRs fed to the channel decoder and the MI between the bits and the extrinsic LLRs output by the channel decoder, respectively. Similar definitions of $I_{A,det}$ and $I_{E,det}$ are used for the SCMA detector. It can be seen from Fig. 4.5 that at $E_b/N_0 = 4dB$, an open tunnel is attained and both curves reach the (1,1) point of perfect convergence to a vanishingly low BER. Hence, it shows that the proposed algorithm is expected to converge.

4.6 Conclusions

In this chapter, an energy minimisation based low-complexity iterative receiver is proposed for SCMA systems. By factorising the joint distribution into the product of several potentials, the Bethe approximation is used to derive the marginals of

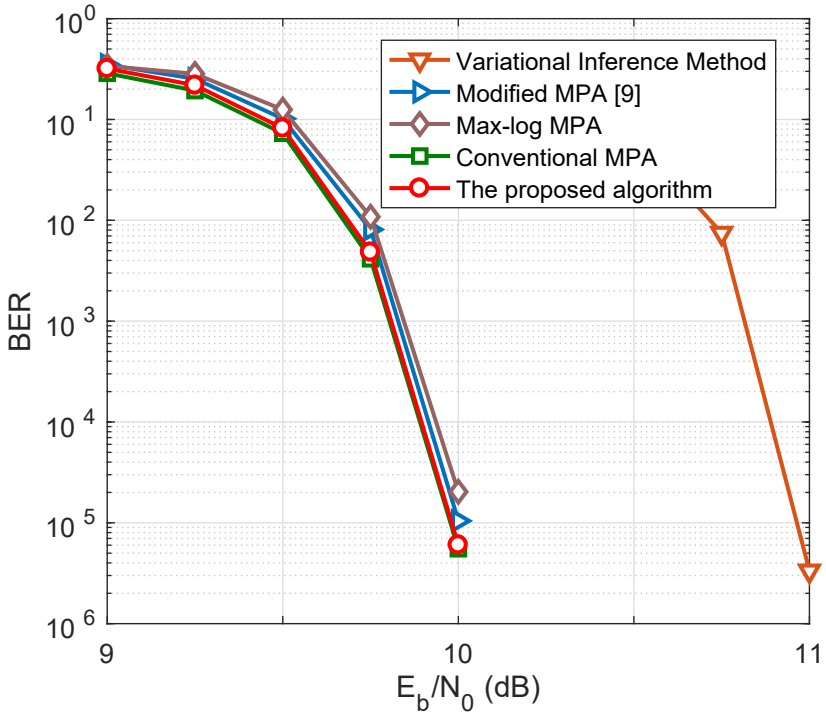


Figure 4.3 : BER performance of different algorithms. ($\lambda = 200\%$)

data symbols. The complexity of the proposed algorithm only increases linearly with the number of users, instead of the exponential complexity of the optimal MAP detector. The convergence of the proposed algorithm is further analysed and its convergence conditions are derived. The simulation results for two SCMA systems with normalised user-load $\lambda = 150\%$ and $\lambda = 200\%$, respectively, show that the performance of the low-complexity energy minimisation based algorithm closely approaches the performance of the conventional MPA scheme and outperforms both the modified MPA and the variational inference methods.

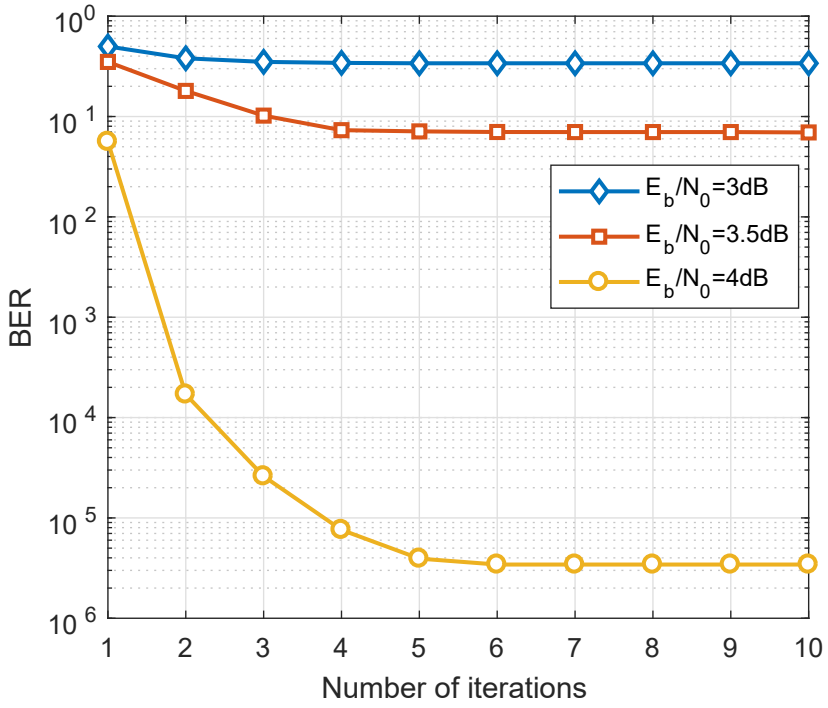


Figure 4.4 : BER performance versus the number of iterations. ($\lambda = 150\%$)

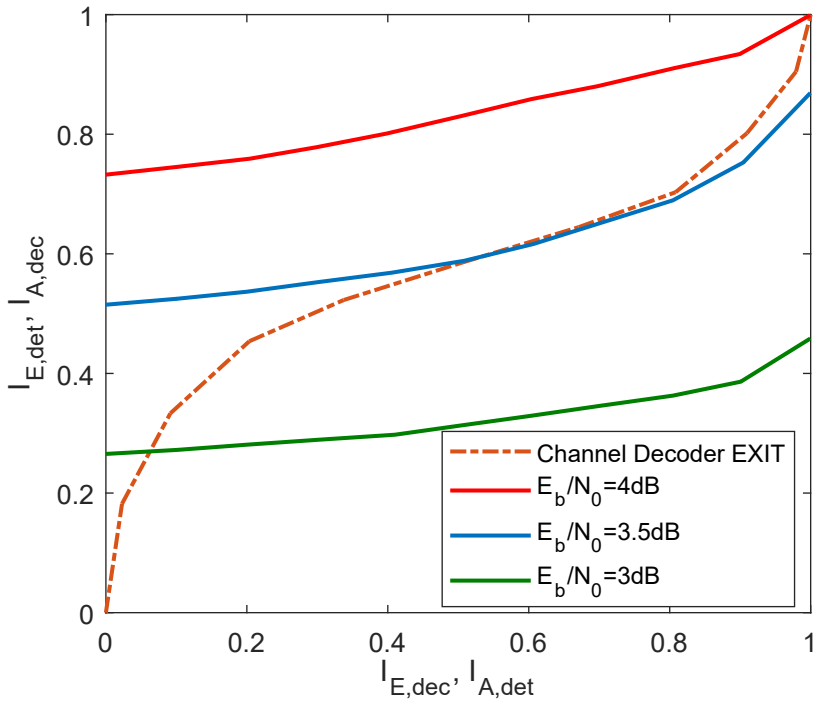


Figure 4.5 : EXIT chart between the SCMA detector and channel decoder. ($\lambda = 150\%$)

Chapter 5

Downlink MIMO-SCMA Receiver Design: Convergent Message Passing and Cooperative Detection

5.1 Introduction

In Chapter 4, a low complexity receiver is designed for uplink SCMA system. This Chapter will focus on receiver design problem of downlink SCMA problem. Considering spatial diversity, SCMA can be combined with the multiple-input multiple-output (MIMO) systems to further improve the spectral efficiency. Different from the orthogonal multiple access, the optimal detection for MIMO-SCMA system suffers from very high computational complexity [105]. In [106], Gaussian distribution is utilised to approximate data symbols and two low-complexity MPA-based detectors were proposed for MIMO-NOMA system. In wideband communications, channel becomes frequency selective and signal suffers from inter-symbol interference (ISI) [100,102]. In [107] and [108], low-complexity receivers based on MPA were proposed for MIMO systems over frequency selective channels. It is well known that the belief propagation (BP) method gives the exact marginal when the factor graph is loop free [94]. However, for MIMO-SCMA system over frequency selective channels, due to the loops of factor graph representations, BP may fail to converge or easily converge to local minima. This phenomenon can be interpreted from the perspective of variational free energy [101]. The BP rules can be derived from minimising constrained Bethe free energy. Unfortunately, when the factor graph contains loops, Bethe free energy is no longer convex and the resultant MPA is not guaranteed

to converge. So far there have been no investigations on the convergent MPA in receiver design problem.

Moreover, for multiuser systems, diversity gain can be achieved by sharing necessary information between users. The users can share its received signal samples between each other directly. However, this kind of mechanism is not realistic due to two reasons: 1) transmitting measurements to a possibly distant user is quite power consuming; 2) a multi-hop routing scheme is required to ensure all measurements are collected, which may not be practical. On the contrary, the distributed processing scheme based on cooperation between users in a network is more attractive since it only requires local computations and communications with neighboring users. The distributed cooperative processing method has been widely used in localisation and target tracking problems [109, 110]. For communications, only a few papers considered the strategies for base station or receiver cooperation but did not apply them to MPA-based receivers in fading channels [111–113].

This chapter aims for designing a convergent version of MPA to be used in factor graph with short cycles and developing cooperative detection schemes in a network with inter-user links. The MIMO-SCMA system with frequency selective fading channels is first introduced. Then to reduce the complexity of the MPA receiver based on factor graph, auxiliary variables are introduced and the factorisation of the joint posterior distribution is represented by a stretched factor graph. Moreover, considering the proposed BP-EP receiver may fail to converge in the MIMO-SCMA scenario due to the loopy factor graph, the Bethe free energy is convexifid and a convergence-guaranteed BP-EP receiver is proposed. Based on the observation that global messages on factor graph can be expressed as the product of several local messages calculated at each user, it is able to perform the cooperative detection in a distributed way. A belief consensus-based scheme is proposed. As all messages are Gaussian distributed, only their means and variances are exchanged between users.

Then considering that the inter-user links are noisy in practice and the convergence speed of belief consensus method will become rather slow, an alternative direction method of multipliers (ADMM)-based algorithm [114] is proposed which aims at minimising the Kullback-Leibler divergence [115] between the global message and the product of local messages.

5.2 Problem Formulation

5.2.1 System Model

A K-user downlink MIMO-SCMA system is considered, where each user is equipped with a single antenna and the base station is equipped with J transmit antennas. The SCMA encoder is a mapping function that maps every $\log_2 M$ coded bits to an J-dimensional SCMA codeword. Let $\mathbf{x}_k^n = [x_{k,1}^n, ..., x_{k,J}^n]^T$ be the transmitted codeword of user k at time instant n. Then the codewords can be multiplexed over J transmit antennas at the base station. The block diagram of the considered system is illustrated in Fig. 5.1.

Let's denote the transmitted symbol at the jth antenna and time instant n by s_j^n, then s_j^n is given by

$$s_j^n = \sum_{k=1}^{K} x_{k,j}^n. \tag{5.1}$$

The signal from base station transmits over frequency selective fading channels with L taps and is received at different users. With the assumption of perfect synchronisation between the base station and users, the received signal at user k and time instant n can be further written as

$$y_k^n = \sum_{j=1}^{J} \sum_{l=0}^{L-1} h_{j,k}^l s_j^{n-l} + \omega_k^n, \tag{5.2}$$

where $h_{j,k}^l$ is the lth tap coefficient of the multipath channel between the jth antenna and kth user, and ω_k^n is additive white Gaussian noise (AWGN) at time instant n

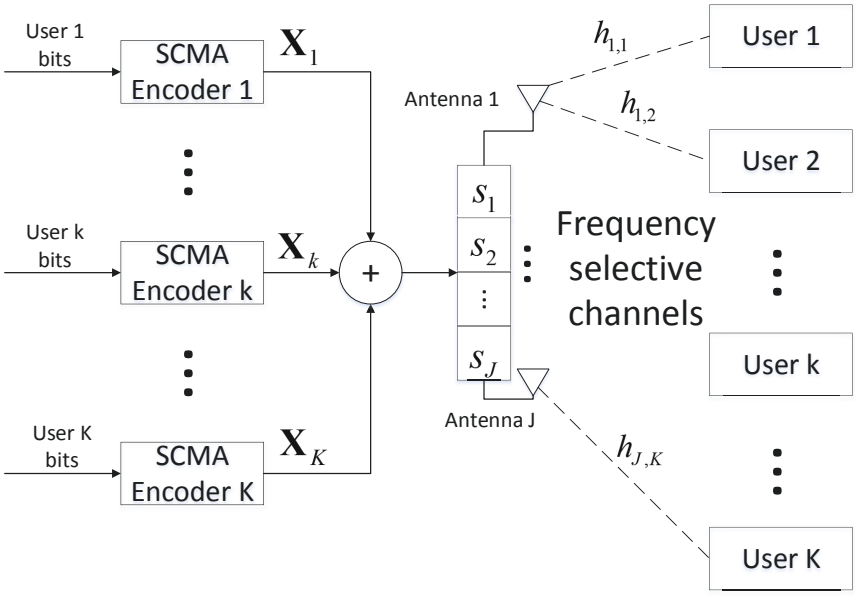


Figure 5.1 : System model for downlink MIMO-SCMA.

with power spectral density N_0.

5.2.2 Probabilistic Model

Assuming that N codewords are transmitted, $\mathbf{X}_k$ and $\mathbf{y}_k$ are denoted as transmitted SCMA codewords and received signal samples of the kth user, and $\mathbf{X}$ as the transmitted symbols of all users. Assuming perfect channel state information, each user can perform the optimal maximum *a posteriori* (MAP) detection based on measurement $\mathbf{y}_k$, which can be expressed as

$$\begin{aligned}\hat{\mathbf{X}}_k &= \arg\max_{\mathbf{X}_k} p(\mathbf{X}_k|\mathbf{y}_k) \\ &= \arg\max_{\mathbf{X}_k} \int p(\mathbf{X}|\mathbf{y}_k)\mathrm{d}\mathbf{X}\backslash\mathbf{X}_k. \end{aligned} \tag{5.3}$$

Following Bayesian rules, $p(\mathbf{X}|\mathbf{y}_k)$ reads

$$p(\mathbf{X}|\mathbf{y}_k) \propto p(\mathbf{X})p(\mathbf{y}_k|\mathbf{X}), \tag{5.4}$$

where $p(\mathbf{X})$ is the joint *a priori* distribution and $p(\mathbf{y}_k|\mathbf{X})$ is the joint likelihood function. Since all transmitted symbols are assumed independent, $p(\mathbf{X}) = \prod_{j,k,n} p(x_{k,j}^n)$,

where $p(x_{k,j}^n)$ is calculated based on the log likelihood ratios (LLRs) of coded bits from the output of channel decoder.

The computational complexity of the optimal MAP receiver in (5.3) increases exponentially with the product of the number of users, the number of antennas and the channel length. In the following, low-complexity message passing receivers are developed for MIMO-SCMA system.

5.3 Low-Complexity BP-EP Receiver based on Stretched Factor Graph

5.3.1 Factor Graph Representation

A factor graph is a bipartite graph representing the factorisation of a function, which enables efficient computations of marginals. Since the noise samples at different time instants are uncorrelated, the joint likelihood function $p(\mathbf{y}_k|\mathbf{X})$ can be factorised as

$$p(\mathbf{y}_k|\mathbf{X}) \propto \prod_n \underbrace{\exp\left(-\frac{\left|y_k^n - \sum_{j=1}^{J}\sum_{l=0}^{L-1} h_{j,k}^l \sum_{k=1}^{K} x_{k,j}^{n-l}\right|^2}{2N_0}\right)}_{f_n^k}. \tag{5.5}$$

Consequently, the joint *a posteriori* distribution can be rewritten as

$$p(\mathbf{X}|\mathbf{y}_k) \propto \prod_n \prod_{j,k} p(x_{k,j}^n) \exp\left(-\frac{\left|y_k^n - \sum_{j=1}^{J}\sum_{l=0}^{L-1} h_{j,k}^l \sum_{k=1}^{K} x_{k,j}^{n-l}\right|^2}{2N_0}\right), \tag{5.6}$$

and can be represented by a factor graph, as depicted in Fig. 5.2, where the shorthand notation f_k^n denotes the local likelihood function corresponding to received sample y_k^n. Squares and circles are used to denote factor vertices and variable vertices. A variable vertex x is connected to a factor vertex f via an edge if and only if x is a variable of the function f.

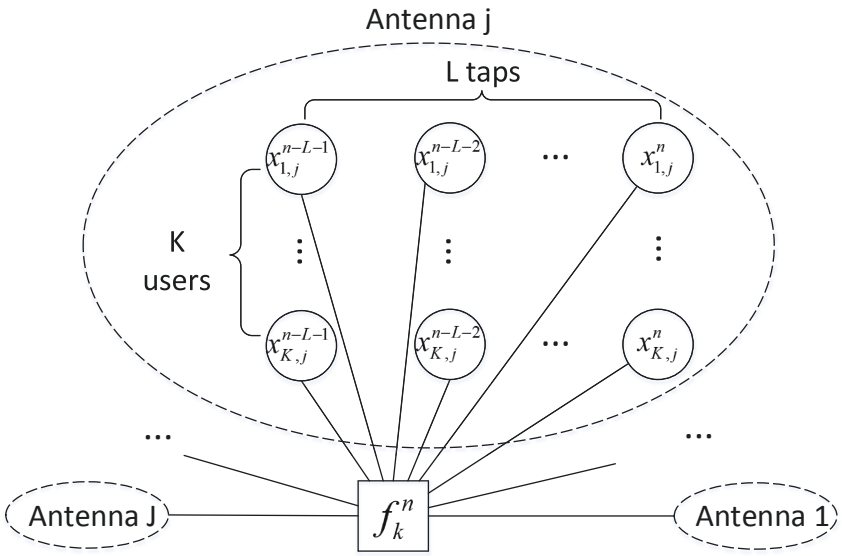


Figure 5.2 : Factor graph representation of the factorisation in (5.6). For ease of exposition, only the variable vertices connected to the factor node f_k^n are plotted.

5.3.2 Stretched Factor Graph and Low-Complexity BP-EP Receiver

The message passing update rules on factor graph has been derived in (1.9) - (1.11). Note that in Fig. 5.2, JKL variables in total are connected to a factor node. Following (1.9), when calculating the message $\mu_{f_k^n \to x_{k,j}^n}(x_{k,j}^n)$, multi-dimensional integration of f_k^n over $(JKL-1)$ variables have to be performed. Therefore, the complexity of the BP receiver based on the factor graph in Fig. 5.2 is $\mathcal{O}(N(JKL)^2)$, which is huge in MIMO-SCMA systems. To tackle this problem, auxiliary variables are introduced to reduce the number of messages that have to be updated, thereby reducing the complexity of receiver significantly. Let $r_{k,j}^n = \sum_{l=0}^{L-1} h_{k,j}^l s_j^{n-l}$, and the likelihood function in (5.5) can be rewritten as

$$p(\mathbf{y}_k|\mathbf{X}) \propto \prod_n \left(\exp\left(-\frac{|y_k^n - \sum_j r_{j,k}^n|^2}{2N_0} \right) \prod_j {}_k^n \phi_j^n \right), \tag{5.7}$$

where $\psi_k^n = \delta(r_{k,j}^n - \sum_{l=0}^{L-1} h_{k,j}^l s_j^{n-l})$ and $\phi_j^n = \delta(s_j^n - \sum_{k=1}^{K} x_{k,j}^n)$ denote the equality constraints. Based on the above factorisation, it is able to stretch multiple variables

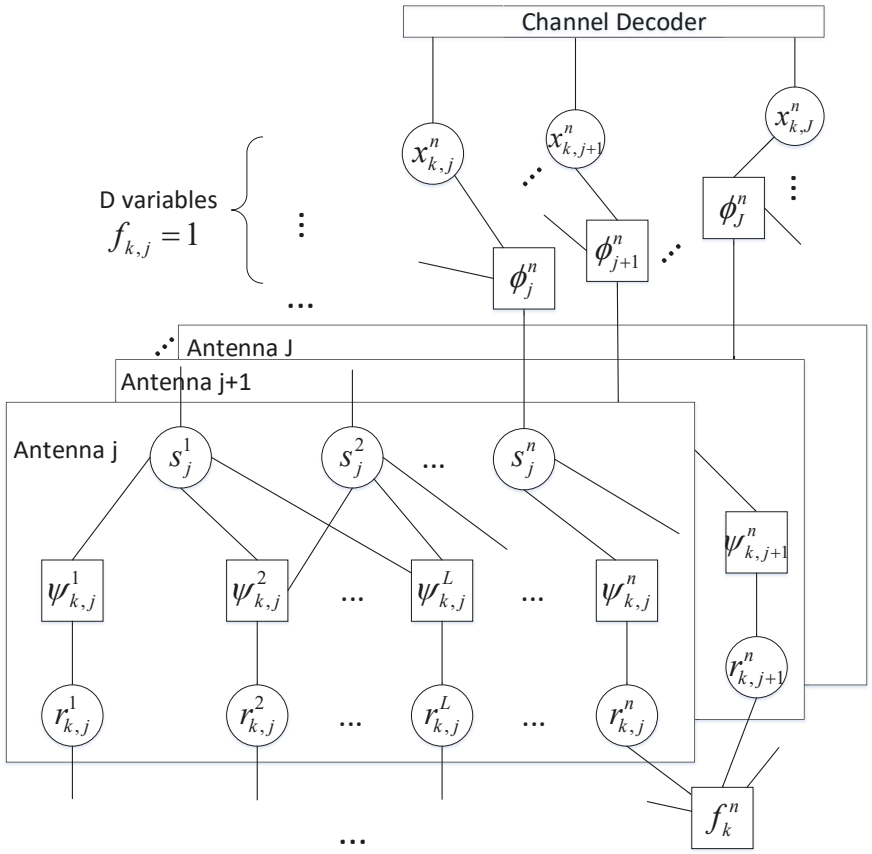


Figure 5.3 : Stretched factor graph representation of the considered MIMO-SCMA system.

and construct a novel factor graph as illustrated in Fig. 5.3, which is named as 'stretched factor graph'.

Based on the rules in (1.9) and (1.10), the messages on factor graph can be updated as follows.

• Incoming Message $\mu_{x^n_{k,j}\to\phi^n_j}(x^n_{k,j})$ (EP updating):

The message $\mu_{x^n_{k,j}\to\phi^n_j}(x^n_{k,j})$ can be regarded as the prior distribution $p(x^n_{k,j})$ of symbols, which is expressed as

$$p(x^n_{k,j}) = \sum_{i=1}^{M} p_i \delta(x^n_{k,j} - \chi_i), \tag{5.8}$$

where χ_i is the ith constellation point, p_i can be calculated based on the LLRs of bits from the output of channel decoder. Generally, $p(x_{k,j}^n)$ can be approximated as Gaussian distribution by directly matching the first and second order moments. However, this will lead to performance loss. To solve this problem, EP is employed which matches the moments of belief instead of the prior distribution [116]. Since information from detector is also exploited, a hybrid BP-EP receiver is expected to improve the performance. Assuming that the message $\mu_{\phi_j^n \to x_{k,j}^n}(x_{k,j}^n)$ can be represented by Gaussian distribution $\mathcal{G}(m_{\phi_j^n \to x_{k,j}^n}, v_{\phi_j^n \to x_{k,j}^n})$, the mean and variance of the belief of $x_{k,j}^n$ can be expressed as

$$m_{x_{k,j}^n} = \frac{1}{2\pi v_{\phi_j^n \to x_{k,j}^n}} \cdot \sum_{i=1}^{M} \chi_i p_i \exp\left(-\frac{(m_{\phi_j^n \to x_{k,j}^n} - \chi_i)^2}{v_{\phi_j^n \to x_{k,j}^n}} \right), \tag{5.9}$$

$$v_{x_{k,j}^n} = \frac{1}{2\pi v_{f \to x_{k,j}^n}} \cdot \sum_{i=1}^{M} |\chi_i|^2 p_i \exp\left(-\frac{(m_{\phi_j^n \to x_{k,j}^n} - \chi_i)^2}{v_{\phi_j^n \to x_{k,j}^n}} \right) - |m_{x_{k,j}^n}|^2. \tag{5.10}$$

Consequently, the Gaussian approximation of message $\mu_{x_{k,j}^n \to \phi_j^n}(x_{k,j}^n)$ is given as

$$m_{x_{k,j}^n \to \phi_j^n} = v_{x_{k,j}^n \to \phi_j^n} \left(\frac{m_{x_{k,j}^n}}{v_{x_{k,j}^n}} - \frac{m_{\phi_j^n \to x_{k,j}^n}}{v_{\phi_j^n \to x_{k,j}^n}} \right), \tag{5.11}$$

$$v_{x_{k,j}^n \to \phi_j^n} = \frac{v_{x_{k,j}^n} v_{\phi_j^n \to x_{k,j}^n}}{v_{\phi_j^n \to x_{k,j}^n} - v_{x_{k,j}^n}}. \tag{5.12}$$

• Messages related to ϕ_j^n and $\quad {}_{k,j}^{n}$ (BP updating):

After collecting the messages $\mu_{x_{k,j}^n \to \phi_j^n}(x_{k,j}^n)$, $\forall k$, and assuming that $\mu_{s_j^n \to \phi_j^n}(s_j^n) \propto \mathcal{G}(m_{s_j^n \to \phi_j^n}, v_{s_j^n \to \phi_j^n})$ is obtained, the message $\mu_{\phi_j^n \to x_{k,j}^n}(x_{k,j}^n)$ is updated according to (1.9), i.e.,

$$\mu_{\phi_j^n \to x_{k,j}^n}(x_{k,j}^n) \propto \int \delta(s_j^n - \sum_{k=1}^{K} x_{k,j}^n) \prod_{k' \neq k} \mu_{x_{k',j}^n \to \phi_j^n}(x_{k',j}^n) \mu_{s_j^n \to \phi_j^n}(s_j^n) \mathrm{d}s_j^n \mathrm{d}x_{k',j}^n$$

$$\propto \mathcal{G}(m_{\phi_j^n \to x_{k,j}^n}, v_{\phi_j^n \to x_{k,j}^n}), \tag{5.13}$$

with the mean and variance given as follows,

$$m_{\phi_j^n \to x_{k,j}^n} = m_{s_j^n \to \phi_j^n} - \sum_{k' \neq k} m_{x_{k',j}^n \to \phi_j^n}, \tag{5.14}$$

$$v_{\phi_j^n \to x_{k,j}^n} = v_{s_j^n \to \phi_j^n} + \sum_{k' \neq k} v_{x_{k',j}^n \to \phi_j^n}. \tag{5.15}$$

The message from ϕ_j^n to s_j^n can also be expressed as Gaussian distribution with mean and variance as

$$m_{\phi_j^n \to s_j^n} = \sum_{k} m_{x_{k,j}^n \to \phi_j^n}, \tag{5.16}$$

$$v_{\phi_j^n \to s_j^n} = \sum_{k \neq k} v_{x_{k,j}^n \to \phi_j^n}. \tag{5.17}$$

Similarly, the means and variances of the messages $\mu_{\psi_{k,j}^n \to s_j^{n-l}}(x_{k,j}^n)$ and $\mu_{\psi_{k,j}^n \to r_{k,j}^n}$ are obtained as

$$m_{\psi_{k,j}^n \to s_j^{n-l}} = \frac{1}{h_{k,j}^l} \left(m_{r_{k,j}^n \to \psi_{k,j}^n} - \sum_{l'=0, l' \neq l}^{L-1} h_{k,j}^{l'} m_{s_j^{n-l'} \to \psi_{k,j}^n} \right), \tag{5.18}$$

$$v_{\psi_{k,j}^n \to s_j^{n-l}} = \frac{v_{r_{k,j}^n \to \psi_{k,j}^n} + \sum_{l'=0, l' \neq l}^{L-1} |h_{k,j}^{l'}|^2 v_{s_j^{n-l'} \to \psi_{k,j}^n}}{|h_{k,j}^l|^2}, \tag{5.19}$$

$$m_{\psi_{k,j}^n \to r_{k,j}^n} = \sum_{l=0}^{L-1} h_{k,j}^l m_{s_j^{n-l} \to \psi_{k,j}^n}, \tag{5.20}$$

$$v_{\psi_{k,j}^n \to r_{k,j}^n} = \sum_{l=0}^{L-1} h_{k,j}^l v_{s_j^{n-l} \to \psi_{k,j}^n}, \tag{5.21}$$

where $m_{s_j^n \to \psi_{k,j}^n}$ and $v_{s_j^n \to \psi_{k,j}^n}$ are given as

$$m_{s_j^n \to \psi_{k,j}^n} = v_{s_j^n \to \psi_{k,j}^n} \left(\frac{m_{\phi_j^n \to s_j^n}}{v_{\phi_j^n \to s_j^n}} + \sum_{l=1}^{L-1} \frac{m_{\psi_{k,j}^{n+l} \to s_j^n}}{v_{\psi_{k,j}^{n+l} \to s_j^n}} \right), \tag{5.22}$$

$$v_{s_j^n \to \psi_{k,j}^n} = \left(v_{\phi_j^n \to s_j^n}^{-1} + \sum_{l=1}^{L-1} v_{\psi_{k,j}^{n+l} \to s_j^n}^{-1} \right)^{-1}. \tag{5.23}$$

• Messages related to $r_{k,j}^n$ (BP updating):

Finally, the message $\mu_{f_k^n \to r_{k,j}^n}(r_{k,j}^n)$ is calculated. As the message $\mu_{\psi_{k,j}^n \to r_{k,j}^n}$ is known, the outgoing message from f_k^n to $r_{k,j}^n$ can be calculated as

$$\mu_{f_k^n \to r_{k,j}^n}(r_{k,j}^n) \propto \int \exp\left(-\frac{|y_k^n - \sum_j r_{k,j}^n|^2}{2N_0}\right) \prod_{j' \neq j} \mu_{r_{k,j'}^n \to f_k^n}(r_{k,j'}^n) \mathrm{d}r_{k,j'}^n, \tag{5.24}$$

which is in Gaussian form with parameters

$$m_{f_k^n \to r_{k,j}^n} = y_k^n - \sum_{j' \neq j} m_{r_{k,j'}^n \to f_k^n}, \tag{5.25}$$

$$v_{f_k^n \to r_{k,j}^n} = N_0 + \sum_{j' \neq j} v_{r_{k,j'}^n \to f_k^n}. \tag{5.26}$$

Remark that by introducing the auxiliary variables, the modified factor graph based on (5.7) is able to reduce the number of integrated messages to $(J+K+L-1)$, which is much lower than $JKL-1$, especially when the number of users and antennas is large.

5.3.3 Algorithm Summary

From the expressions in Section III.B, updating the messages relies on other variables. Therefore, the messages on the factor graph are updated iteratively. At the first iteration, since no information about the transmitted data symbols is given, the messages $\mu_{g_{k,j}^n \to x_{k,j}^n}(x_{k,j}^n), \forall k, j, n$ are initialised as zero mean Gaussian distribution. Then in each iteration, the mean and variance values of all messages are updated according to the rules derived as in (5.11)-(5.26). After determining the message $\mu_{\phi_j^n \to x_{k,j}^n}(x_{k,j}^n)$, the detector calculates the extrinsic information of bits and feeds them to the channel decoder. After decoding, the channel decoder updates the LLRs of bits and starts the next iteration. The details of the proposed stretched factor graph-base BP-EP receiver is summarised in Algorithm 3.

Algorithm 3 Stretched Factor Graph-based BP-EP Receiver for MIMO-SCMA System over Frequency Selective Channels

Initialisation:

The incoming messages are initialised as zero mean Gaussian distribution with zero mean and infinite variance;

for iter=1:N_{Iter} **do**

    Compute the means and variances of downward messages according to (5.16), (5.17) and (5.20)-(5.23);

    Compute the message from factor vertex f_k^n to variable vertex $r_{k,j}^n$ according to (5.25)-(5.26);

    Compute the means and variances of upward messages according to (5.14), (5.15) and (5.18), (5.19);

    Convert the outgoing messages to LLR and feed them to the channel decoder;

    Perform standard BP channel decoding;

    Calculate the incoming messages using EP as in (5.11) and (5.12);

end for

5.4 Convergence-guaranteed BP-EP Receiver

The main drawback of the above algorithm based on standard BP is that it does not guarantee convergence in loopy graphs, e.g., the factor graph illustrated in Fig. 5.3. Several papers have investigated this issue, e.g., tree reweighted BP and dampening message method. In this section, how to obtain belief propagation message updating rules based on the variational free energy framework [94] is first introduced. Then an iterative message passing receiver with guaranteed convergence is proposed by convexifying the Bethe free energy.

5.4.1 Variational Free Energy and Belief Propagation

Consider a joint distribution $p(\mathbf{x})$ of random variables $\mathbf{x} = [x_1, ..., x_i, ...]$ that can be factorised into the product of several non-negative functions as

$$p(\mathbf{x}) = \prod_a f_a(\mathbf{x}_a), \tag{5.27}$$

where a is the index of function f_a with arguments $\mathbf{x}_a$ and $\mathbf{x}_a \triangleq (x_i | i \in \mathcal{N}(a))$. The factorisation in (5.27) can be represented by a factor graph. Calculating the marginal distribution $p(x_i) = \sum_{\mathbf{x} \setminus x_i} p(\mathbf{x})$ requires the summation over the states of all variables except x_i. The variational method is an efficient way to find approximate solutions for the marginals.

Let $b(\mathbf{x})$ be a positive function approximating $p(\mathbf{x})$, then the variational free energy is defined as the Kullback-Leibler divergence between $b(\mathbf{x})$ and $p(\mathbf{x})$ [115], i.e.,

$$\begin{aligned} F = \mathbb{D}[b(\mathbf{x})|p(\mathbf{x})] &= \int b(\mathbf{x}) \ln \frac{b(\mathbf{x})}{p(\mathbf{x})} \mathrm{d}\mathbf{x} \\ &= \underbrace{\int b(\mathbf{x}) \ln b(\mathbf{x}) \mathrm{d}\mathbf{x}}_{-H(b)} - \int b(\mathbf{x}) \ln p(\mathbf{x}) \mathrm{d}\mathbf{x}, \end{aligned} \tag{5.28}$$

where $H(b)$ is the entropy of $b(\mathbf{x})$. The goal is to find $b(\mathbf{x})$ which minimises the variational free energy. To keep consistency with the form of $p(\mathbf{x})$, the Bethe approximation [117] is employed, given by

$$b(\mathbf{x}) = \frac{\prod_a b_a(\mathbf{x}_a)}{\prod_i b_i(x_i)^{|\mathcal{N}(i)|-1}}. \tag{5.29}$$

In (5.29), $b_a(\mathbf{x}_a)$ is the joint belief of variables $\mathbf{x}_a$ and $b_i(x_i)$ is the approximate marginal of x_i.

Substituting (5.29) into (5.28), the Bethe free energy is obtained as

$$\begin{aligned} F_B = &- \sum_a \int b_a(\mathbf{x}_a) \ln f_a(\mathbf{x}_a) \mathrm{d}\mathbf{x}_a + \sum_a \int b_a(\mathbf{x}_a) \ln b_a(\mathbf{x}_a) \mathrm{d}\mathbf{x}_a \\ &+ \sum_i (1 - |\mathcal{N}(i)|) \int b_i(x_i) \ln b_i(x_i) \mathrm{d}x_i. \end{aligned} \tag{5.30}$$

Considering the normalisation constraint and marginalisation constraint, the corresponding Lagrangian is constructed as

$$\begin{aligned}\mathcal{L}_B \triangleq & F_B + \sum_i \beta_i \left(\int b_i(x_i) \mathrm{d}x_i - 1 \right) + \sum_a \beta_a \left(\int b_a(\mathbf{x}_a) \mathrm{d}\mathbf{x}_a - 1 \right) \\ & + \sum_i \sum_{f_a \in \mathcal{N}(i)} \int \beta_{ai}(x_i) \left(b_i(x_i) - \int b_a(\mathbf{x}_a) \mathrm{d}\mathbf{x}_a \backslash x_i \right) \mathrm{d}\mathbf{x}_i. \end{aligned} \tag{5.31}$$

Setting the partial derivatives of $\mathcal{L}_B$ with respect to β_i, β_a and β_{ai} to zero result in the normalisation constraint and marginalisation constraint. Setting $\nabla_{b_a(\mathbf{x}_a)} = 0$ gives

$$\ln b_a(\mathbf{x}_a) = \ln f_a(\mathbf{x}_a) + \sum_{i \in \mathcal{N}(a)} \beta_{ai}(x_i) + \beta_a - 1. \tag{5.32}$$

Similarly, setting $\nabla_{b_i(x_i)} = 0$ can obtain

$$(|\mathcal{N}(i)| - 1) \ln b_i(x_i) = 1 - \beta_i + \sum_{f_a \in \mathcal{N}(i)} \beta_{ai}(x_i). \tag{5.33}$$

Taking exponential function on both sides of (5.32) and (5.33) and setting $\exp(\beta_{ai}(x_i)) = \prod_{f_{a'} \in \mathcal{N}(i) \backslash f_a} \mu_{f_{a'} \to i}(x_i)$ yield

$$b_a(\mathbf{x}_a) \propto f_a(\mathbf{x}_a) \prod_{i \in \mathcal{N}(a)} \prod_{f_{a'} \in \mathcal{N}(i) \backslash f_a} \mu_{f_{a'} \to i}(x_i), \tag{5.34}$$

$$b_i(x_i) \propto \prod_{f_a \in \mathcal{N}(i)} \mu_{f_a \to i}(x_i). \tag{5.35}$$

According to the marginalisation constraint, integrating all variables in $b_a(\mathbf{x}_a)$ except x_i gives

$$\begin{aligned} b_i(x_i) & \propto \int b_a(\mathbf{x}_a) \mathrm{d}\mathbf{x}_a \backslash x_i \\ & \propto \int f_a(\mathbf{x}_a) \prod_{i' \in \mathcal{N}(a) \backslash i} \prod_{f_{a'} \in \mathcal{N}(i') \backslash f_a} \mu_{f_{a'} \to i'}(x_{i'}) \mathrm{d}x_{i'} \prod_{f_{a'} \in \mathcal{N}(i) \backslash f_a} \mu_{f_{a'} \to i}(x_i). \end{aligned} \tag{5.36}$$

By comparing (5.36) with (5.35), the message $\mu_{f_a \to x_i}(x_i)$ is expressed as

$$\mu_{f_a \to x_i}(x_i) \propto \int f_a(\mathbf{x}_a) \prod_{i' \in \mathcal{N}(a) \backslash i} \prod_{f_{a'} \in \mathcal{N}(i') \backslash f_a} \mu_{f_{a'} \to x_{i'}}(x_{i'}) \mathrm{d}\mathbf{x}_a \backslash x_i, \tag{5.37}$$

which is the same as the message updating rule in (1.9). Define

$$\mu_{x_i \to f_a}(x_i) = \frac{b_i(x_i)}{\mu_{f_a \to x_i}(x_i)} \propto \prod_{f_{a'} \in \mathcal{N}(i) \backslash f_a} \mu_{f_{a'} \to x_i}(x_i), \tag{5.38}$$

which is the rule in (1.10).

5.4.2 Convergence-guaranteed BP-EP Receiver

In previous subsection, it is shown that the message updating rules of the standard BP can be derived by minimising the constrained Bethe free energy. However, in general, the Bethe free energy is non-convex and has several local minima. As a result, BP is not guaranteed to converge.

Using the definition $H_i(b) = -\int b_i(x_i) \ln b_i(x_i) \mathrm{d}x_i$ and $H_a(b) = -\int b_a(\mathbf{x}_a) \ln b_a(\mathbf{x}_a) \mathrm{d}\mathbf{x}_a$, the entropy $H(b)$ is rewritten as $H(b) = \sum_i (1 - \mathcal{N}(i)) H_i(b) + \sum_a H_a(b)$. As shown by Yedidia [94], the Bethe approximation for entropy can be generalised by linearly combining $H_i(b)$ and $H_a(b)$ as

$$\tilde{H}(b) = \sum_i c_i H_i(b) + \sum_a c_a H_a(b), \tag{5.39}$$

where the counting numbers c_i and c_a satisfy $c_i = 1 - \sum_{f_a \in \mathcal{N}(i)} c_a$. Obviously, $c_a = 1$ corresponds to the Bethe approximation. Consequently, the constrained optimisation problem in (5.31) can be revised as

$$\begin{aligned} \min \ & -\sum_a \int b_a(\mathbf{x}_a) \ln f_a(\mathbf{x}_a) \mathrm{d}\mathbf{x}_a - \sum_a c_a H_a(b) - \sum_i c_i H_i(b) \\ \mathtt{s.t}\ & b_i(x_i) = \int b_a(\mathbf{x}_a) \mathrm{d}\mathbf{x}_a \backslash x_i, \int b_a(\mathbf{x}_a) \mathrm{d}\mathbf{x}_a = 1, \int b_i(x_i) \mathrm{d}\mathbf{x}_i = 1. \end{aligned} \tag{5.40}$$

Based on (5.40), a Lagrangian is constructed and corresponding message updating rules are derived. Similarly, the notations in (1.9) and (1.10) are used to denote

the messages, which are given as

$$\mu_{f_a \to x_i}(x_i) = \left(\tilde{\mu}_{f_a \to x_i}(x_i)\right)^{\gamma_{ai}} \left(\tilde{\mu}_{x_i \to f_a}(x_i)\right)^{\gamma_{ia}-1}, \tag{5.41}$$

$$\mu_{x_i \to f_a}(x_i) = \left(\tilde{\mu}_{f_a \to x_i}(x_i)\right)^{\gamma_{ai}-1} \left(\tilde{\mu}_{x_i \to f_a}(x_i)\right)^{\gamma_{ia}}, \tag{5.42}$$

where $\gamma_{ai} = |\mathcal{N}(i)|c_a/(|\mathcal{N}(i)|c_a + c_i + |\mathcal{N}(i)| - 1)$ and $\gamma_{ia} = |\mathcal{N}(i)|/(|\mathcal{N}(i)|c_a + c_i + |\mathcal{N}(i)| - 1)$. The messages $\tilde{\mu}_{f_a \to x_i}(x_i)$ and $\tilde{\mu}_{x_i \to f_a}(x_i)$ are given as

$$\tilde{\mu}_{f_a \to x_i}(x_i) \propto \int f_a^{\frac{1}{c_a}}(\mathbf{x}_a) \prod_{i' \in \mathcal{N}(a) \backslash i} \mu_{x_{i'} \to f_a}(x_{i'}) \mathrm{d}\mathbf{x}_a \backslash x_i, \tag{5.43}$$

$$\tilde{\mu}_{x_i \to f_a}(x_i) \propto \prod_{f_{a'} \in \mathcal{N}(i) \backslash f_a} \mu_{f_{a'} \to i}(x_i). \tag{5.44}$$

The detailed derivation of messages (5.41) and (5.42) is given in Appendix. E. Note that in (5.41)-(5.44), if $c_a = 1$ and $c_i = 1 - |\mathcal{N}(i)|$, the message updating rules become the standard BP.

Based on the general form of Bethe free energy, it is possible to find appropriate counting numbers to form a convex free energy. The prominent tree reweighted BP approximates the free energy as the combination of several entropy terms over spanning trees [118]. Then the edge appearance probability is used as counting number c_a to convexify the free energy. Nevertheless, only a few convex free energies can be represented using spanning trees. To this end, a more general condition is considered, as stated in the following proposition.

Proposition 3:

The modified Bethe free energy is convergence-guaranteed if there exist non-negative counting numbers c_{ia}, c_{ii} and c_{aa} satisfying

$$c_a = c_{aa} + \sum_{i \in \mathcal{N}(a)} c_{ia}, \tag{5.45}$$

$$c_i = c_{ii} - \sum_{f_a \in \mathcal{N}(i)} c_{ia}. \tag{5.46}$$

Proof:

Substituting (5.45) and (5.46) into (5.30) yields

$$F_B = -\sum_a \int b_a(\mathbf{x}_a) \ln f_a(\mathbf{x}_a) \mathrm{d}\mathbf{x}_a - \sum_a c_{aa} H_a(b) - \sum_i c_{ii} H_i(b) - \sum_{i, f_a \in \mathcal{N}(i)} c_{ia}(H_a(b) - H_b(b)). \tag{5.47}$$

For the first term on the right hand side of (5.47), the second order partial derivatives with respect to $b_a(\mathbf{x}_a)$ equals 0. Therefore the convexity of F_B is dominated by the convexity of

$$F_{conv} = -\sum_a c_{aa} H_a(b) - \sum_i c_{ii} H_i(b) - \sum_{i, f_a \in \mathcal{N}(i)} c_{ia}(H_a(b) - H_i(b)). \tag{5.48}$$

Since $\frac{\partial^2 H_i(b)}{\partial b_i(x_i)^2} = -\frac{1}{b_i(x_i)}$ and $\frac{\partial^2 H_a(b)}{\partial b_a(\mathbf{x}_a)^2} = -\frac{1}{b_a(\mathbf{x}_a)}$ hold, F_{conv} is convex if and only if $c_{aa} \geq 0$, $c_{ii} \geq 0$ and $-\sum_{i, f_a \in \mathcal{N}(i)} c_{ia}(H_a(b) - H_i(b))$ is convex.

To analyse the convexity of $H_a(b) - H_i(b)$, the idea from [119] is used that $b_i(x_i) = \int b_a(\mathbf{x}_a) d\mathbf{x}_a \backslash x_i = b_a(x_i)$. Then $F_{ai} = H_i(b) - H_a(b)$ can be written as

$$F_{ai} = \int b_a(\mathbf{x}_a) \ln b_a(\mathbf{x}_a) \mathrm{d}\mathbf{x}_a - \int b_a(x_i) \ln b_i(x_i) \mathrm{d}x_i. \tag{5.49}$$

The second order partial derivatives of F_{ai} with respect to $b_a(\mathbf{x}_a)$ and $b_i(\mathbf{x}_i)$ are expressed as

$$\frac{\partial^2 F_{ai}}{\partial b_a(x_a)^2} = \frac{1}{b_a(\mathbf{x}_a)}, \tag{5.50}$$

$$\frac{\partial^2 F_{ai}}{\partial b_i(x_i)^2} = -\frac{b_a(x_i)}{(b_i(x_i))^2}, \tag{5.51}$$

$$\frac{\partial^2 F_{ai}}{\partial b_a(\mathbf{x}_a) \partial b_i(x_i)} = \frac{\partial^2 F_{ai}}{\partial b_i(x_i) \partial b_a(\mathbf{x}_a)} = -\frac{1}{b_i(x_i)}. \tag{5.52}$$

Convexity of F_{ai} is satisfied if the Hessian matrix with components (5.50)-(5.52) is

positive semidefinite. For any beliefs $\tilde{b} = [\tilde{b}_a(\mathbf{x}_a), \tilde{b}_i(x_i)]$,

$$\begin{aligned}
&\int \tilde{b} \begin{bmatrix} \frac{\partial^2 F_{ai}}{\partial b_a(x_a)^2} & \frac{\partial^2 F_{ai}}{\partial b_a(\mathbf{x}_a)\partial b_i(x_i)} \\ \frac{\partial^2 F_{ai}}{\partial b_i(x_i)\partial b_a(\mathbf{x}_a)} & \frac{\partial^2 F_{ai}}{\partial b_i(x_i)^2} \end{bmatrix} \tilde{b}^T \mathrm{d}\mathbf{x}_a \\
&= \int \left(\frac{(\tilde{b}_a(\mathbf{x}_a))^2}{b_a(\mathbf{x}_a)} - \frac{2\tilde{b}_a(\mathbf{x}_a)\tilde{b}_i(x_i)}{\tilde{b}_i(x_i)} + \frac{\tilde{b}_a(x_i)(\tilde{b}_i(x_i))^2}{(b_i(x_i))^2} \right) \mathrm{d}\mathbf{x}_a \\
&= \int \tilde{b}_a(\mathbf{x}_a) \left(\frac{(\tilde{b}_a(\mathbf{x}_a))}{b_a(\mathbf{x}_a)} - \frac{\tilde{b}_i(x_i)}{b_i(x_i)} \right)^2 \mathrm{d}\mathbf{x}_a \geq 0, \qquad (5.53)
\end{aligned}$$

which indicates that F_{ai} is convex. Therefore, under the conditions (5.45) and (5.46), the modified Bethe free energy is convergence-guaranteed. ■

From the above equations, various counting numbers can be chosen to obtain different approximate free energy. Since the goal is deriving a convergence-guaranteed version of BP algorithm, the modified free energy should be close to the Bethe free energy. Denoting the counting numbers of Bethe approximation as $d_a = 1$ and $d_i = 1 - |\mathcal{N}(i)|$, minimising the l_2 norm $\|\mathbf{c} - \mathbf{d}\|^2$ can be formulated as a quadratic program as

$$\begin{aligned}
&\min_{c_{ii}, c_{aa}, c_{ia}} \sum_a \left(c_{aa} + \sum_{i \in \mathcal{N}(a)} c_{ia} - 1 \right)^2 \qquad (5.54) \\
&\texttt{s.t} \;\; c_i = 1 - \sum_{f_a \in \mathcal{N}(i)} c_a, \;\; (5.45), (5.46).
\end{aligned}$$

The above optimisation problem can be easily solved using standard solvers and therefore will not be elaborated here.

It is well known that for real numbers a and b, $(e^a)^b = e^{ab}$. Therefore, if $\tilde{\mu}_{f_a \to x_i}(x_i)$ is obtained in Gaussian, $(\tilde{\mu}_{f_a \to x_i}(x_i))^{\gamma_{ai}}$ is still Gaussian with the same mean and variance divided by γ_{ai}. Then similar to Section 5.3.2, Gaussian messages can be derived based on modified message passing rules with c_i and c_a. The details of the proposed convergence-guaranteed message passing algorithm is summarised in Algorithm 4.

Algorithm 4 Convergence-guaranteed BP-EP Receiver for MIMO-SCMA System over Frequency Selective Channels

Initialisation:
The incoming messages are initialised as Gaussian distribution with zero mean and infinite variance;
Calculate the counting numbers by solving (5.54);
for iter=1:N_{Iter} **do**
    Compute the auxiliary messages from factor vertices to variable vertices according to (5.43);
    Compute the auxiliary messages from variable vertices to factor vertices according to (5.44);
    Compute the messages $\mu_{f_a \to x_i}$ and $\mu_{x_i \to f_a}$ using (5.42) and (5.41);
    Convert the outgoing messages to LLR and feed them to the channel decoder;
    Perform standard BP channel decoding;
    Calculate the incoming messages using expectation propagation as in (5.11) and (5.12);
end for

5.4.3 Complexity

The complexities of proposed BP-EP receiver based on original and stretched factor graphs have already been analysed in Section III. B. The complexity of proposed convergence-guaranteed BP-EP receiver is also dominated by the number of messages to be integrated. As the number of integrated messages does not change, the complexity required for message passing is still $\mathcal{O}(N(J+K+L-1)^2)$. Note that in (5.54), a quadratic programming should be solved to determine the counting number, which improves the number of operations. Nevertheless, since (5.54) can be done off-line before perform detection, the complexity of proposed convergence-

guaranteed BP-EP receiver is $\mathcal{O}(N(J+K+L-1)^2)$. In Table I, the computational complexities of different receivers are compared.

Table 5.1 : Computational Complexities of Different Receivers

Receivers	Computational Complexity
MPA	$\mathcal{O}(N \cdot 2^{JKL-1})$
BP-EP (Original Factor Graph)	$\mathcal{O}(N(JKL-1)^2)$
BP-EP (Stretched Factor Graph)	$\mathcal{O}(N(J+K+L-1)^2)$
Convergence-guaranteed BP-EP	$\mathcal{O}(N(J+K+L-1)^2)$

5.5 Distributed Cooperative Detection

The MIMO-SCMA system in cooperative network is further considered where users are allowed to communicate with each other. Since the transmitted symbols s_j^n from base station are received by all the users, cooperative detection can be performed to exploit the diversity gain.

Thanks to the factor graph framework, it is possible to represent the relationship of s_j^n and measurements observed at different users graphically, as illustrated in Fig. 5.4. Note that for each user, the stretched factor graph proposed in Section 5.3 is adopted and the convergence-guaranteed BP-EP algorithm proposed in Section 5.4 is employed. Denoting $\mu_k(s_j^n)$ as the message to variable vertex s_j^n based on the measurement of user k, then

$$\mu_k(s_j^n) = \prod_{l=0}^{L-1} \mu_{\psi_k^{n+l} \to s_j^n}(s_j^n). \tag{5.55}$$

Having the messages $\mu_k(s_j^n),\ \forall k$, the message $\mu_{s_j^n \to \phi_j^n}(s_j^n)$ and the extrinsic information corresponding to data symbol $x_{k,j}^n$ are calculated. Obviously, if the measure-

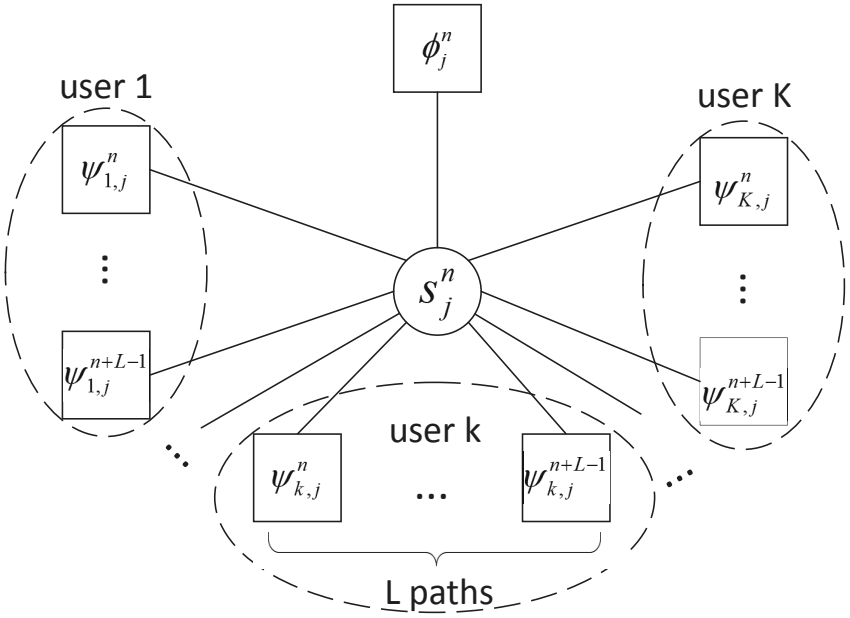


Figure 5.4 : Factor graph representation for cooperative detection.

ments are collected by a central unit, it is easy to obtain $\mu_{s_j^n \to \phi_j^n}(s_j^n)$ as the product of $\mu_k(s_j^n)$, $\forall k$. However, transmitting measurements to a possibly distant central unit leads to huge power consumption. On the contrary, using distributed method only requires local communications with neighboring users. By exchanging packets between neighboring users, all users can fully exploit the information related to s_j^n. In what follows, two distributed cooperative detection methods for the considered MIMO-SCMA system are devised.

5.5.1 Belief Consensus-Based Method

Let $\mathcal{S}_k$ denotes the neighboring set of user k, i.e., every user in $\mathcal{S}_k$ can communicate with user k. The goal is to determine $\mu_{s_j^n \to \phi_j^n}(s_j^n)$ ('global message') at each user with only local processing and communications. The belief consensus method is efficient to compute the product of several local functions over the same variable distributively.

With its local message $\mu_k(s_j^n)$ based on measurement $\mathbf{y}_k$, user k updates its local belief $\rho_k^{p+1}(s_j^n)$ according to standard belief consensus recursion, i.e.,

$$\rho_k^{p+1}(s_j^n) = \rho_k^p(s_j^n) \prod_{i \in \mathcal{S}_k} \left(\frac{\rho_i^p(s_j^n)}{\rho_k^p(s_j^n)} \right)^{\eta}, \tag{5.56}$$

where the superscript p denotes the index of consensus iterations and η is the update rate. At the first iteration, the local belief is initialised as $\rho_k^0(s_j^n) = \mu_k(s_j^n)$. In the standard belief consensus, all users share the a constant updating rate η, which may cause performance loss. The metropolis weight [120] can be used to solve this problem as

$$\rho_k^{p+1}(s_j^n) = \rho_k^p(s_j^n)^{\eta_{kk}} \prod_{i \in \mathcal{S}_k} \rho_i^p(s_j^n)^{\eta_{ki}}, \tag{5.57}$$

with the update rate

$$\eta_{ik} = \eta_{ki} = \begin{cases} 1/\max(|\mathcal{S}_k|, |\mathcal{S}_i|), & \text{for } i \neq k \\ 1 - \sum_{i' \in \mathcal{S}_k} \eta_{i'k}, & \text{for } i = k. \end{cases} \tag{5.58}$$

With the assumption of Gaussian messages, users are able to exchange parameters of the messages instead of the distribution. In this case, (5.57) can be rewritten as

$$\boldsymbol{\theta}_k^{p+1} = \eta_{kk} \boldsymbol{\theta}_k^p + \sum_{i \in \mathcal{S}_k} \eta_{ki} \boldsymbol{\theta}_i^p, \tag{5.59}$$

where $\boldsymbol{\theta}_k^p = [m_{k \to s_j^n}^p / v_{k \to s_j^n}^p, 1/v_{k \to s_j^n}^p]^T$ represents the parameters to be exchanged. Consequently, users only broadcast $\boldsymbol{\theta}_k^p$ to neighboring users.[1] For a connected graph that each user has at least one neighbor, after running several consensus iterations, all users reach consensus on the global message, i.e., $\rho_k^{N_p}(s_j^n) = \mu_{s_j^n \to \phi_j^n}(s_j^n)^{1/K}$,

[1]It may happen that in a consensus iteration, the link between two users, e.g., user i and k fails. In this case, an additional variable $\bar{\boldsymbol{\theta}}_{ik}$ is used to store the parameter received in previous iteration. Then $\bar{\boldsymbol{\theta}}_{ik}$ can be used to continue consensus updating.

$\forall k$ [110]. However, when exchanging packets between users, inter-user links may suffer from additive noise, which causes the variance of $\boldsymbol{\theta}_k$ growing unbounded. To tackle this problem, a vanishing parameter is introduced and (5.59) is replaced by

$$\boldsymbol{\theta}_k^{p+1} = \boldsymbol{\theta}_k^p + \alpha^p \sum_{i \in \mathcal{S}_k} \eta_{ki} \left(\boldsymbol{\theta}_i^p + \boldsymbol{\omega}_{ki}^p - \boldsymbol{\theta}_k^p \right), \tag{5.60}$$

where $\boldsymbol{\omega}_{ki}$ is the additive noise on link $i \to k$ and α^p is the vanishing parameter. However, since α^p decreases monotonically as the increase of p, the convergence speed of (5.60) will be rather slow.

5.5.2 Bregeman ADMM-Based Method

To reach the consensus, the product of all local beliefs should be as close as possible to the global message. Motivated by this, the distributed processing problem can be reformulated as the minimisation of the Kullback-Leibler divergence with the constraint $\rho_k(s_j^n) = \rho_i(s_j^n)$:

$$\begin{aligned} \min_{\boldsymbol{\rho}} \quad & \mathbb{D}[\mu_{s_j^n \to \phi_j^n}(s_j^n) | \prod_k \rho_k(s_j^n)] \\ \texttt{s.t} \quad & \rho_k(s_j^n) = \rho_i(s_j^n), \ \forall \ k, i \in \mathcal{S}_k. \end{aligned} \tag{5.61}$$

Considering that $\rho_k(s_j^n)$ can be characterised by $\boldsymbol{\theta}_k$, the objective functional can be minimised subject to $\boldsymbol{\theta}_k = \boldsymbol{\theta}_i$. For decoupling purpose, a set of additional variables $\boldsymbol{\pi}_{ki}$ are defined for each inter-user link, through which the optimisation problem can be rewritten as

$$\begin{aligned} \min_{\boldsymbol{\theta}} \quad & \mathbb{D}[\mu_{s_j^n \to \phi_j^n}(s_j^n) | \prod_k \rho_k(s_j^n)] \\ \texttt{s.t} \quad & \boldsymbol{\theta}_k = \boldsymbol{\pi}_k, \ \boldsymbol{\theta}_i = \boldsymbol{\pi}_k, \ \forall k, i \in \mathcal{S}_k. \end{aligned} \tag{5.62}$$

The optimisation problem in (5.62) subject to equality constraints can be solved by ADMM. In each iteration, ADMM updates variables in a block coordinate fashion

by solving the augmented Lagrangian of (5.62), which is defined as

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{\lambda}) = \mathbb{D}[\mu_{s_j^n \to \phi_j^n}(s_j^n) | \prod_k \rho_k(s_j^n)] + \sum_k \sum_{i \in \mathcal{S}_k} \left(\boldsymbol{\lambda}_{kk}^T (\boldsymbol{\theta}_k - \boldsymbol{\pi}_k) + \boldsymbol{\lambda}_{ki}^T (\boldsymbol{\theta}_i - \boldsymbol{\pi}_k) \right) + \frac{c}{2} \sum_k \sum_{i \in \mathcal{S}_k} \left(\|\boldsymbol{\theta}_k - \boldsymbol{\pi}_k\|_2^2 + \|\boldsymbol{\theta}_i - \boldsymbol{\pi}_k\|_2^2 \right), \tag{5.63}$$

where $c > 0$ denotes the penalty coefficient and $\boldsymbol{\lambda}$ is the associated Lagrangian multipliers. The quadratic penalty term may result in a high complexity when solving (5.63). To this end, the quadratic penalty term is replaced by Bregman divergence to generalise ADMM. Let ε be a continuously differentiable and strictly convex function, namely, Bregman function. The Bregman divergence is defined as

$$B_\varepsilon(x, y) = \varepsilon(x) - \varepsilon(y) - \langle x - y, \nabla_\varepsilon(y) \rangle, \tag{5.64}$$

where $\nabla_\varepsilon(y)$ is the gradient of ε and $\langle \cdot \rangle$ denotes the inner product.

Based on Bregeman divergence, the following augmented Lagrangian is given,

$$\mathcal{L}_{Breg}(\boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{\lambda}) = \mathbb{D}[\mu_{s_j^n \to \phi_j^n}(s_j^n) | \prod_k \rho_k(s_j^n)] + \sum_k \sum_{i \in \mathcal{S}_k} \left(\boldsymbol{\lambda}_{ki}^T (\boldsymbol{\theta}_k - \boldsymbol{\pi}_k) + \boldsymbol{\lambda}_{ik}^T (\boldsymbol{\pi}_k - \boldsymbol{\theta}_i) \right) + c \sum_k \sum_{i \in \mathcal{S}_k} B_\varepsilon(\boldsymbol{\theta}_i, \boldsymbol{\pi}_k). \tag{5.65}$$

Following the Bregman ADMM, $\mathcal{L}_B$ is minimised with respect to one set of variables given the others. At the $(p+1)$th iteration, the updates for Bregman ADMM can be described as

$$\boldsymbol{\theta}^{p+1} = \arg\min_{\boldsymbol{\theta}} \mathcal{L}_{Breg}(\boldsymbol{\theta}, \boldsymbol{\pi}^p, \boldsymbol{\lambda}^p), \tag{5.66}$$

$$\boldsymbol{\pi}^{p+1} = \arg\min_{\boldsymbol{\pi}} \mathcal{L}_{Breg}(\boldsymbol{\theta}^{p+1}, \boldsymbol{\pi}, \boldsymbol{\lambda}^p), \tag{5.67}$$

$$\boldsymbol{\lambda}_{ki}^{p+1} = \boldsymbol{\lambda}_{ki}^p + c(\boldsymbol{\theta}_i^{p+1} - \boldsymbol{\pi}_{ki}^{p+1}). \tag{5.68}$$

Actually, a series of ε can be chosen to form different Bregman divergences. For efficient computations, an appropriate Bregman function is to be given. In this case, since the messages are in Gaussian form, the log partition function is employed as

Bregman function. Consequently, the Bregman divergence between two variables is equivalent to the Kullback-Leibler divergence between two Gaussian distributions characterised by these two variables, i.e., $B_\varepsilon(\mathbf{a},\mathbf{b}) = \mathbb{D}[f(\mathbf{x}|\mathbf{a})|f(\mathbf{x}|\mathbf{b})]$. Then (5.66)-(5.68) can be analytically derived as

$$\boldsymbol{\theta}_k^{p+1} = \frac{\boldsymbol{\theta}_k^0 + \sum_{i\in\mathcal{S}_k\cup k}\left(\boldsymbol{\lambda}_{ki}^p + c\boldsymbol{\pi}_i^p\right)}{1 + c(|\mathcal{S}_k|+1)}, \tag{5.69}$$

$$\boldsymbol{\pi}_k^{p+1} = \frac{\sum_{i\in\mathcal{S}_k\cup k}\left(c\boldsymbol{\theta}_i^{p+1} - \boldsymbol{\lambda}_{ki}^p\right)}{c(|\mathcal{S}_k|+1)}, \tag{5.70}$$

$$\boldsymbol{\lambda}_{ki}^{p+1} = \boldsymbol{\lambda}_{ki}^p + c(\boldsymbol{\theta}_k^{p+1} - \boldsymbol{\pi}_i^{p+1}). \tag{5.71}$$

By exchanging $\boldsymbol{\theta}_k^p$ and $\boldsymbol{\pi}_k^p$ in the network, all users can obtain the message $\mu_{s_j^n\to\phi_j^n}(s_j^n)$ in a distributed fashion. As the penalty parameter c will affect the convergence speed, c are assumed to be different for different users which is varying in each iteration [121], given as

$$c_k^{p+1} = \begin{cases} c_k^p\cdot(1+\tau) & \text{if} \|\boldsymbol{\epsilon}_k^t\|_2 > \kappa\|\boldsymbol{\iota}_k^t\|_2 \\ c_k^p\cdot(1+\tau)^{-1} & \text{if} \|\boldsymbol{\iota}_k^t\|_2 > \kappa\|\boldsymbol{\epsilon}_k^t\|_2 \\ c_k^p & \text{otherwise,} \end{cases} \tag{5.72}$$

where $\|\boldsymbol{\epsilon}_k^t\|_2$ and $\|\boldsymbol{\iota}_k^t\|_2$ are the primal and dual residuals, defined as $\|\boldsymbol{\epsilon}_k^t\|_2 = \|\boldsymbol{\theta}_k^p - \bar{\boldsymbol{\theta}}_k^p\|_2$, $\|\boldsymbol{\iota}_k^t\|_2 = \|\bar{\boldsymbol{\theta}_k}^p - \bar{\boldsymbol{\theta}}_i^{p-1}\|_2$, $\bar{\boldsymbol{\theta}_k}^p = \frac{1}{|\mathcal{S}_k|}\sum_{i\in\mathcal{S}_k}\boldsymbol{\theta}_i^p$. Typical values of κ and τ are suggested as constant $\kappa = 10$ and $\tau = 1$.

Propostion 4:

Following Bregman ADMM updates rules, all local parameters can reach consensus on the global parameter.

Proof:

Note that for all k, the second order partial derivative of the functional $\mathbb{D}[\mu_{s_j^n\to\phi_j^n}(s_j^n)|\prod_k\rho_k(s_j^n)]$ satisfies

$$\frac{\partial^2\mathbb{D}[\mu_{s_j^n\to\phi_j^n}(s_j^n)|\prod_k\rho_k(s_j^n)]}{\partial\rho_k(s_j^n)^2} = \frac{1}{\rho_k(s_j^n)} > 0,$$

which shows the optimisation objective is convex. Moreover, since ε is strictly convex, the Bregman penalty term is also convex. This implies that the optimisation problem to be solved is convex and the convergence is guaranteed. ■

After several iterations, the local belief $\rho_k^{N_p}(s_j^n)$, $\forall k$ characterised by $\boldsymbol{\theta}_k^{N_c}$ is guaranteed to converge to the global message $\mu_{s_j^n \to \phi_j^n}(s_j^n)$. Compared to belief consensus-based method, it can be seen the Bregman ADMM-based algorithm requires to transmit an additional variable, which doubles the communication overhead.

If the inter-user links suffer from the additive noise, the updates (5.69) and (5.71) can be interpreted as stochastic gradient updates, whose variances have been proved to be bounded values [122].

5.5.3 Algorithm Summary

For the distributed cooperative detection, it is assumed that each user has obtained the message $\mu_{k \to s_j^n}(s_j^n), \forall k, j, n$ based on its local measurements, and the goal is to obtain the product of all users' messages distributively. To start with the distributed algorithm, the local belief $\rho_k^0(s_j^n)$ is initialised as $\mu_{k \to s_j^n}(s_j^n)$. Then according to belief consensus-based method and Bregman ADMM-based method, all users update its local belief to reach agreement on the global message. Also, with the advantage of Gaussian distribution, only few parameters are exchanged in cooperative detection. For both schemes, the complexity is $\mathcal{O}(N)$, which linearly increases with the number of users, making them attractive in practical applications. The proposed distributed cooperative detection methods are summarised in Algorithm 5.

5.6 Simulation Results

The performance of the proposed receivers is evaluated by Monte Carlo simulations and compared with several state-of-the-art methods. A MIMO-SCMA system

Algorithm 5 Belief Consensus and Bregman ADMM-based Methods for Distributed Cooperative Detection

1: Each user computes message $\mu_{k\to s_j^n}(s_j^n), \forall k, j, n$ based on its local measurements.

2**Enter** cooperative detection

3: Initialise $\rho_k^0(s_j^n)$ as $\mu_{k\to s_j^n}(s_j^n)$

4: **for** p=1:N_p **do**

5: Each user broadcasts the parameters $\boldsymbol{\theta}_k^p$ (*Belief Consensus*) or $\boldsymbol{\theta}_k^p, \boldsymbol{\pi}_k^p, \forall k$ (*Bregman ADMM*) to its neighboring users;

6: Each user update its local parameters using (5.59) (*Belief Consensus*) or (5.69)- (5.71) (*Bregman ADMM*) ;

7: **end for**

8: Calculate the message $\mu_{s_j^n\to\phi_j^n}(s_j^n)$ at all receivers;

9:**Exit** cooperative detection

10: Computes other messages on factor graph with $\mu_{s_j^n\to\phi_j^n}(s_j^n)$.

with $J = 4$ antennas, $K = 6$ users, $D = 2$ nonzero entries in each codeword and $M = 4$ is considered. Therefore, the overloading factor is $\varrho = 150\%$. The SCMA codebook is designed according to [39] with the indicator matrix $\mathbf{F}$ defined as

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}. \tag{5.73}$$

A 5/7-rate LDPC code is employed with variable and check node degree distributions being $v(X) = 0.0005 + 0.2852X + 0.2857X^2 + 0.4286X^3$ and $c(X) = 0.0017X^9 + 0.9983X^{10}$. Quadrature phase shifting key (QPSK) is utilised as the mother modulation scheme. The channel is frequency selective with $L = 10$ taps, and

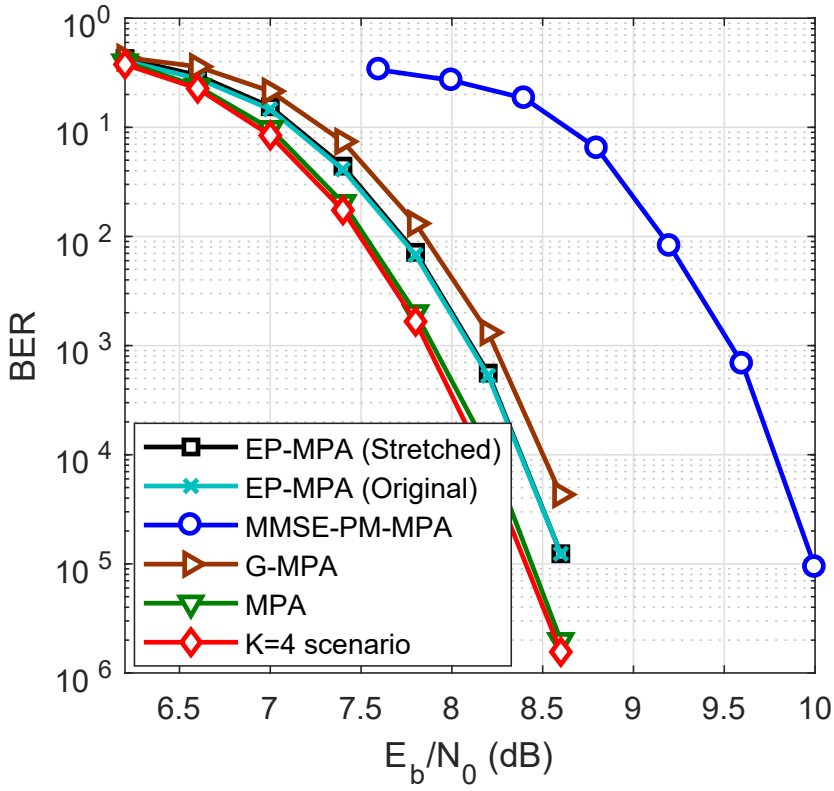


Figure 5.5 : BER performance of different algorithms for MIMO-SCMA system.

each tap is independently generated according to the distribution $h_{k,j}^{l} \sim \mathcal{N}(0, q^{l})$, $\forall k, j$. The normalised power delay profile is $q^{l} = \frac{\exp(-0.1l)}{\sum q^{l}}$. The maximum number of iterations is $N_{Iter} = 10$. The simulation results are averaged from 1000 independent Monte Carlo trails.

In Fig. 5.5, the bit error rate (BER) performance of the proposed stretched factor graph-based BP-EP algorithm (denoted as 'Stretch-BP-EP') in Section III.B is plotted. For comparison, the performance of the MPA receiver [123], Gaussian approximated BP (denoted as 'GaussAppro-BP') algorithm and a combined MMSE-PM-MPA algorithm are also included[2]. A $K = 4$ scenario in which information

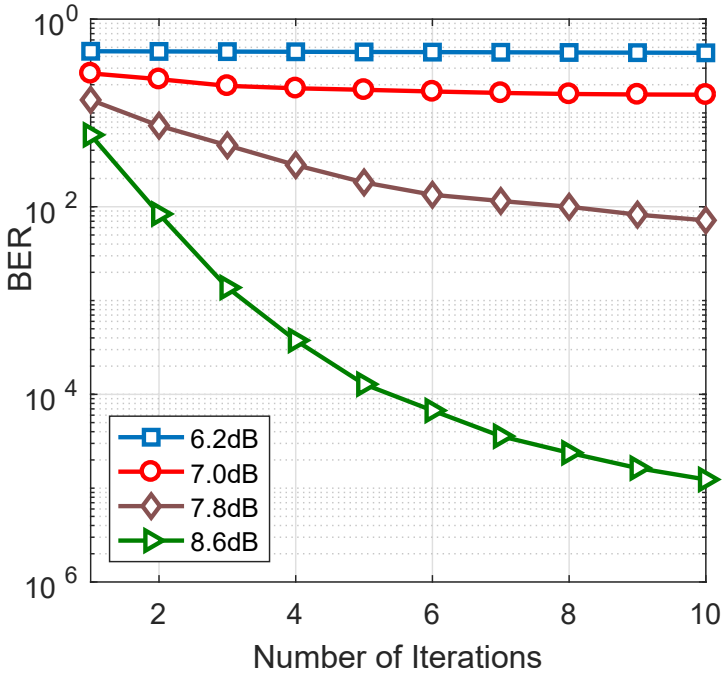
[2]Gaussian approximated BP is also based on the proposed stretched factor graph. However, the extrinsic information of data symbols are approximated by Gaussian directly, instead of approximating the belief as that in EP. The combined MMSE-PM-MPA receiver first performs the

of different users is transmitted using different antennas is also considered as the performance bound (the coding and modulation scheme are assumed to be identical to SCMA). It is observed from Fig. 5.5 that MMSE-PM-MPA method suffers from significant performance loss. This is because that MMSE detector can only output hard information for the PM-MPA-based SCMA detector. The proposed Stretch-BP-EP algorithm slightly outperforms GaussAppro-BP and performs close to the MPA receiver. However, compared to MPA receiver with exponential complexity with respect to the number of interfered numbers, using EP and Gaussian approximation, the complexity of the proposed algorithm is reduced significantly. Moreover, the proposed SCMA system has similar performance compared with the $K = 4$ scenario, while the former is able to support 50% more users.

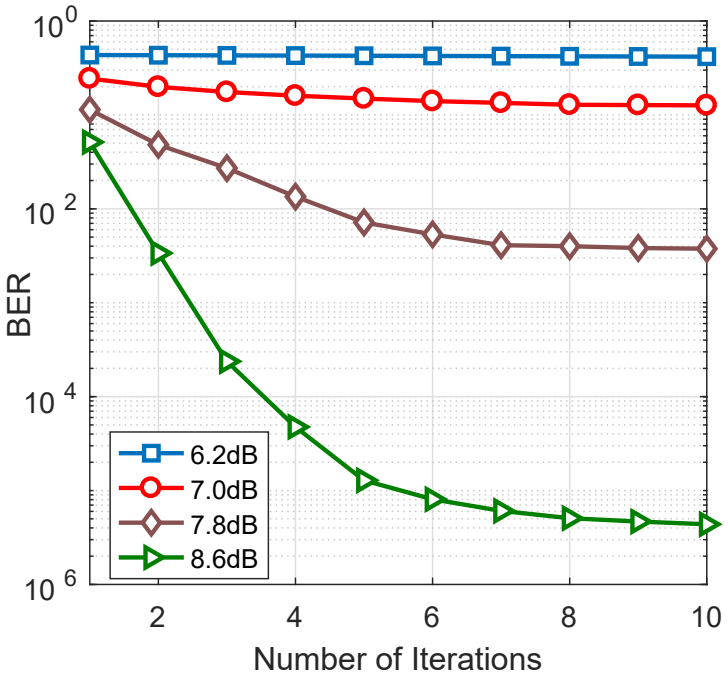
BER performance of the Stretch-BP-EP method and the proposed convergence-guaranteed BP-EP (denoted as 'Conv-BP-EP') algorithm are compared in Fig. 5.6 at different values of E_b/N_0. It is seen that performance of both algorithms improve as the number of iteration increases. After several iterations, the performance gain of both methods become marginal. By comparing Fig. 5.6(a) with Fig. 5.6(b), it can be observed that Conv-BP-EP algorithm converges faster than the Stretch-BP-EP method. This results can be explained by the fact that Stretch-BP-EP algorithm may converge to the local minima of variational free energy while Conv-BP-EP is guaranteed to converge to the global minimum, which demonstrates the superiority of the proposed Conv-BP-EP method.

In the following, the performance of the proposed distributed cooperative detection schemes is evaluated. Six users are uniformly distributed on a $20 \times 20m^2$ unit square. Two users can communicate and exchange information if and only if their distance is less than $d = 10m$. The channels between users are modeled as AWGN

MMSE-based MIMO equalisation and then use PM-MPA [60] for SCMA decoding.



(a) Stretched factor graph-based BP-EP algorithm (Stretch-BP-EP)



(b) Convergence-guaranteed BP-EP algorithm (Conv-BP-EP)

Figure 5.6 : Impact of the number of iterations on BER performance of Stretch-BP-EP and Conv-BP-EP algorithms.

and set to be the same for all links. The vanishing parameter for belief consensus is set to a typical value $\alpha^p = \frac{1}{p}$.

In Fig. 5.7, performance of the proposed two distributed cooperative detection schemes with perfect inter-user links are evaluated. As a benchmark, the BER performance of a centralised scheme is also plotted[3]. Two values of the number of consensus iterations are considered, i.e. $p = 5$ and $p = 10$. For comparison, the averaged BER performance of all users based on their local measurements as in Fig. 5.6(b) is also included. It is observed that, by performing cooperative detection, BER performance can be significantly improved, which reveals that diversity gain can be achieved by exchanging information between neighboring users. By comparing the belief consensus-based method and the Bregman ADMM-based method, it can be seen that, with perfect inter-user links assumption, both methods deliver similar BER performance at $p = 5$ and $p = 10$. Moreover, after 10 iterations, both methods attain the performance of centralised processing.

Since the maximum communication range of inter-user link is critical to the power consumption of users, the BER performance of Bregman ADMM-based algorithm is compared with different communication ranges $d = 2m$, $d = 6m$, $d = 10m$, $d = 14m$ and $d = 20m$. Obviously, increasing d will result in more neighboring users and higher probability of being a fully connected network. As shown in Fig. 5.8, BER performance improves as d increases. However, the performance gain becomes smaller when d is large enough. Considering that the power consumption will increase exponentially as d increases, the power cost and BER performance can be compromised in practice.

The performance of the proposed distributed cooperative detection algorithms

[3]Remark that only the measurements from connected users are collected at a central unit for fair comparison.

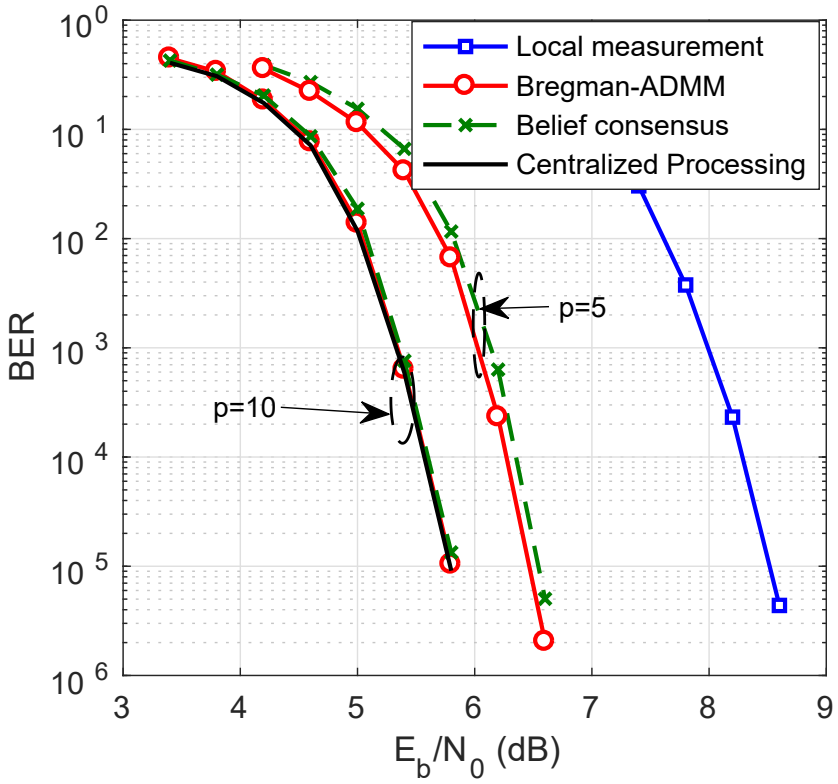


Figure 5.7 : BER performance of the proposed distributed cooperative detection schemes with $p = 5$ and $p = 10$.

are further evaluated in the condition of noisy inter-user links. In Fig. 5.9, the BER performance of the proposed distributed algorithms versus E_b/N_0 is plotted, where the SNR corresponding to the inter-user links are set to be 10dB. The number of consensus iterations is $p = 10$. It is seen that due to the noisy inter-user links, the performance of both distributed schemes at $p = 10$ cannot attain that of the centralised scheme. It is also observed that the Bregman ADMM-based method outperforms the belief consensus-based algorithm. To further analyse the convergence properties of the two distributed schemes, in Fig. 5.10, the mean squared error (MSE) of local parameters $\boldsymbol{\theta}_k$ versus the number of consensus iterations is illustrated. The MSE is defined as $\sum_{k=1}^{K} \|\boldsymbol{\theta}_k - \bar{\boldsymbol{\theta}}\|_2^2$ where $\bar{\boldsymbol{\theta}} = \frac{\sum_k \boldsymbol{\theta}_k}{K}$. Three SNR scenarios of the inter-user links are considered, i.e., SNR $= \{5, 10, 20\}$dB. As ex-

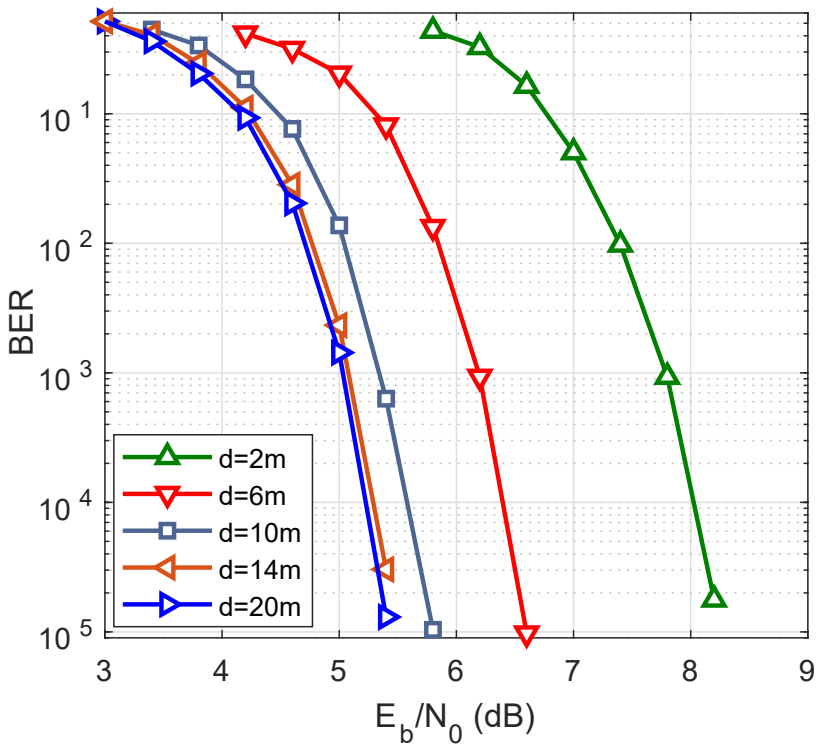


Figure 5.8 : Impact of communication range on the BER performance of Bregman ADMM-based method.

pected, larger SNR leads to smaller uncertainty and the MSE performance is better. Due to the vanishing factor $\alpha^p = \frac{1}{p}$, belief consensus-based algorithm converges slower than the Bregman ADMM-based method. Note that the MSE performance gap between them becomes even larger at higher SNR. This is due to the fact that belief consensus algorithm uses the same vanishing factor at high SNR while Bregman ADMM-based method benefits from small noise variance. Therefore, Bregman ADMM is more efficient in noisy inter-user link networks.

5.7 Conclusions

In this chapter, a graph-based low-complexity message passing receiver for MIMO-SCMA system over frequency selective channels is proposed. Since the direct factorisation of the joint posterior distribution leads to huge complexity in message updat-

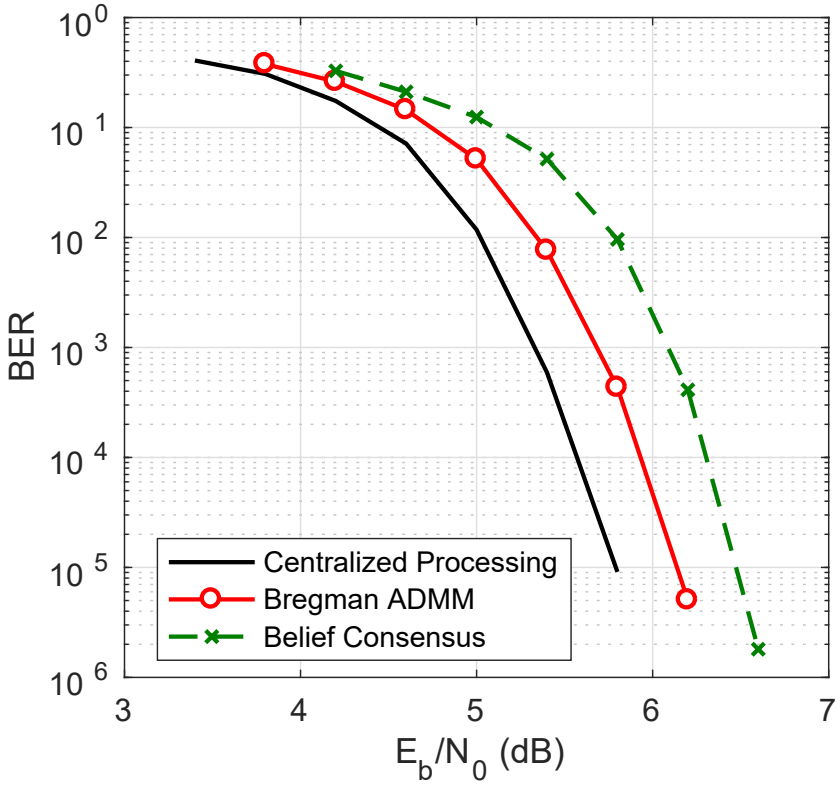


Figure 5.9 : BER performance of the proposed distributed cooperative detection schemes with noisy inter-user links.

ing, auxiliary variables are introduced and a stretched factor graph is constructed. EP is employed to approximate the messages of data symbols to Gaussian distribution and a hybrid BP-EP receiver is proposed.Considering the poor convergence property of the standard BP on loopy factor graph, appropriate counting numbers are used to convexify the Bethe free energy and convergence-guaranteed BP-EP receiver is derived.A cooperative network is further considered and two distributed cooperative detection schemes, i.e., belief consensus-based algorithm and Bregman ADMM-based method, are proposed. The proposed iterative receivers are evaluated by Monte Carlo simulations and compared with the other schemes. It is shown that the proposed Stretch-BP-EP receiver performed close to the MMSE-based receiver with much lower complexity. The proposed Conv-BP-EP receiver outperforms the

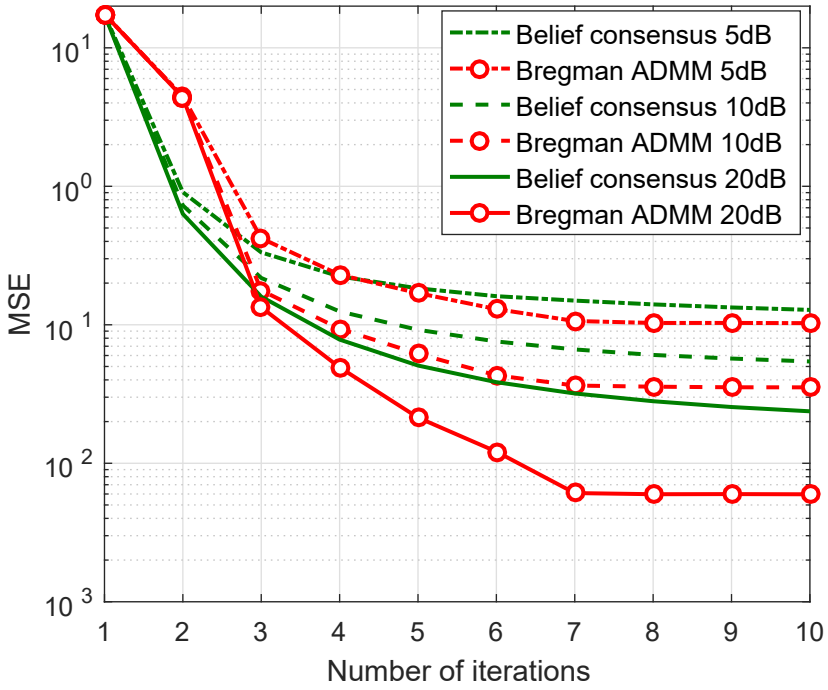


Figure 5.10 : MSE of parameters versus the number of consensus iterations.

Stretch-BP-EP by improving the convergence property. Compared with the orthogonal multiple access counterpart, MIMO-SCMA system with the proposed receivers is shown to be able to support 50% more users over frequency selective fading channels, with negligible BER performance loss. In cooperative networks, it is verified that BER performance can be further improved by exploiting the diversity gain using the proposed two distributed cooperative detection schemes. Moreover, compared with the belief consensus-based algorithm, Bregman ADMM-based method is shown to be more attractive in practical noisy inter-user links.

Chapter 6

Iterative Receiver Design for FTN Signaling - SCMA System

6.1 Introduction

In the above 4 chapters, the receiver design problems for FTN signaling and SCMA system are investigated. Naturally, it is expected to achieve even higher spectral efficiency by using a combined uplink FTN-SCMA system that the data symbols corresponding to a user are further packed using FTN signaling. Nevertheless, the detection of data symbols for an FTN-SCMA system is challenging due to the interferences imposed by non-orthogonal waveforms and non-orthogonal multiple access.

In this chapter, the low complexity joint channel estimation and decoding algorithm based on factor graph and message passing algorithm is studied for an uplink FTN-SCMA systems. To tackle the colored noise induced by FTN signaling, the auto regressive (AR) process is employed to model the noise. Then the joint distribution of data symbols, channel taps and noise samples can be factorised into several local functions and represented by a factor graph. Even with factor graph, conventional MPA is still intractable to be implemented due to exponential order of complexity. To this end, expectation propagation (EP) method is used which restricts the message from channel decoder as Gaussian distribution. Compared to direct approximation via moment matching, EP method aims to minimise a specified relative entropy between the true marginal and the trail distribution [124]. In EP, the extrinsic information to channel decoder is also considered in the approximation,

which can enhance the BER performance. However since the modulus of channel coefficient does not equal to 1, the Gaussian form of messages is unavailable. To tackle this problem, commencing from the variational framework, a modified factor node is reformed, then variational message passing can be employed. Correspondingly, only means and variances needs to be updated iteratively and the complexity order of the proposed receiver for FTN-SCMA systems only scales linearly with the number of users.

Moreover, it has been shown even in busy hours, only a small percentage of users in wireless networks are active [125]. In current OMA uplink scenarios, a request-grant procedure is used that the base station (BS) schedule the uplink transmission after receiving the request from users [126]. This procedure leads to a large communication overhead, especially for the massive connectivity scenario with a huge number of devices. Therefore the uplink grant-free transmission scheme is highly expected to significantly reduce both communication overhead and transmission latency [32]. In grant-free transmission, the active users directly send signals to the BS without grants. In order to decode information bits from users that are accessed simultaneously, BS has to detect user activity based on the received signal. Motivated by the sparsity of active users, compressive sensing (CS) based multiuser detection method was proposed in [127]. Considering channel estimation, a two stage algorithm which detects user activity using CS first and then perform channel estimation and detection was proposed in [128]. An AMP-expectation maximisation (EM) was proposed in [129] which solves the active user detection and channel estimation problem jointly. In [130] and [131], the authors used the precision parameter of channel coefficient variable to describe the user activity and constructed a factor graph to perform joint detection and channel estimation. Different from [130] and [131], in this chapter, a binary variable is used to represent the active/inactive users. By formulating corresponding factor graph, a modified message passing algorithm is

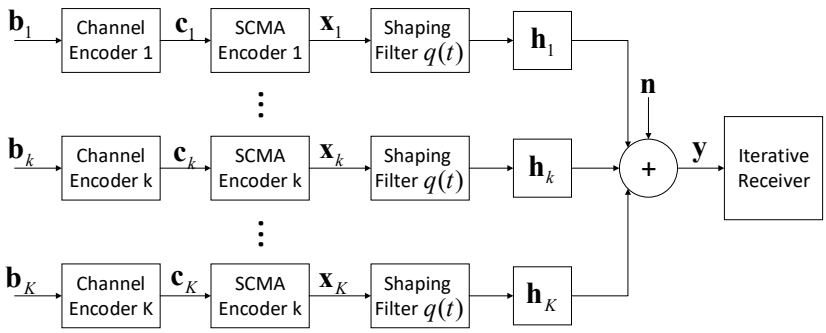


Figure 6.1 : Transmitter side of the considered FTN-SCMA system.

proposed to iteratively calculate the distribution of active users. In addition, to further lower the receiver complexity, EP is used to approximate the binary variable by Gaussian. Accordingly, the proposed receiver still experience low complexity.

6.2 System Model

An uplink SCMA system with K users and J resource elements is considered in this chapter. For brevity the codeword of user k at time instant n is denoted as $\mathbf{x}_k^n = [x_{k1}^n, ..., x_{kJ}^n]^T$. After SCMA encoding, the SCMA codewords are passed through a shaping filter $q(t)$ with symbol period $T = \tau T_0$, where T_0 is the symbol interval of the Nyquist signaling and τ is the FTN packing factor. The modulated signal corresponding to user k over the jth resource elements is formulated as

$$s_{kj}(t) = \sum_n x_{kj}^n q(t - n\tau T_0). \tag{6.1}$$

In Nyquist signaling, $\tau = 1$ guarantees inter symbol interference (ISI) free transmission. In FTN signaling, choosing $0 < \tau < 1$ can transmit more data symbols in the same time period at the cost of introducing intentional ISI. Then the signal corresponding to user k transmit through channel $\mathbf{h}_k = [h_{k1}, ..., h_{kJ}]^T$. The block diagram of the transmit side is shown in Fig. 6.1.

Assuming perfect synchronisation between the users and the base station, the

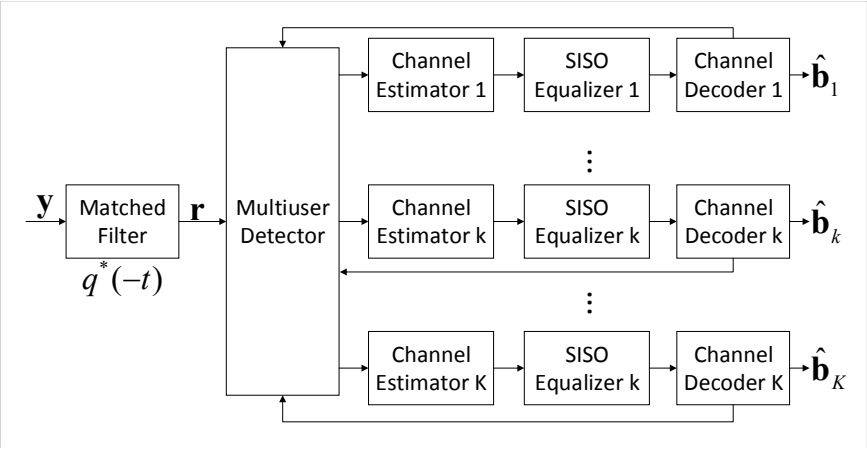


Figure 6.2 : Receiver structure of the considered FTN-SCMA system.

received signal at the base station can be expressed as,

$$\mathbf{y}(t) = \sum_{k=1}^{K} \mathbf{h}_k \odot \mathbf{s}_j(t) + \mathbf{n}(t), \tag{6.2}$$

where $\mathbf{s}_j(t) = [s_{1j}(t), ..., s_{Kj}(t)]^T$ is the modulated signals of all users transmits over the jth resource and $\mathbf{n}_t$ is the additive white Gaussian noise with power spectral density N_0. As shown by Fig. 6.2, the received signal is filtered by a matched filter $q^*(-t)$. Without loss of generality, let $g(t) = q(t) * q*(-t)$. Then the signal is given by

$$\mathbf{r}(t) = \sum_{k=1}^{K} \mathbf{h}_k \odot \sum_{n} x_{kj}^n g(t - n\tau T_0) + \boldsymbol{\omega}(t). \tag{6.3}$$

After sampling with rate $1/\tau T_0$, the samples at the nth time slot is expressed as

$$\mathbf{r}^n = \sum_{k=1}^{K} \mathbf{h}_k \odot \tilde{\mathbf{s}}_k^n + \boldsymbol{\omega}_n, \tag{6.4}$$

where the jth entry in $\tilde{\mathbf{s}}_k^n$ is given as[1]

$$\tilde{s}_{kj}^n = \sum_{i=-L}^{L} g_i x_{kj}^{n-i}, \tag{6.5}$$

[1]In theory, the number of ISI taps induced by FTN is infinite. However in practice, it is possible to choose sufficiently large number of taps, i.e. $2L + 1$ taps.

and $g_{n-i} = \int q(t - n\tau T_0)q^*(t - i\tau T_0)\mathrm{d}t$. In (6.4), ω_n denotes noise samples with respect to all resource elements at time n, formulating as $\omega_n = \int \mathbf{n}(t)q^*(t - n\tau T_0)$. Since the signal rate is above the Nyquist rate, the autocorrelation function of noise sample $\omega_j^n, \forall\ j$ is

$$\mathbb{E}[\omega_j^n \omega_j^m] = N_0 g_{n-m}, \tag{6.6}$$

which indicates in FTN system, the noise at the receiver side is colored noise. To avoid the whitening process that increase the receiver complexity, in the following section, wan autoregressive model aided factor graph approach will be proposed to overcome the colored noise and perform channel estimation and decoding.

6.3 Joint Channel Estimation and Decoding Algorithm for FTN-SCMA Systems

6.3.1 Approximation of Colored Noise

In the same way as shown in Chapter 2, the colored noise is approximated by an P-order AR model as

$$\omega_j^n = \sum_{p=1}^{P} a_p \omega_j^{n-p} + \delta_j^n. \tag{6.7}$$

6.3.2 Probabilistic Model and Factor Graph Representation

Assuming that each user transmits total N SCMA codewords and N samples are received at the base stationthe goal is to determine the *a posteriori* distribution (marinal) of transmitted symbol x_{kj}^n based on all observations at the base station $\mathbf{r}$. Then such marginal is transformed into extrinsic log likelihood ratio (LLR) and fed to the channel decoder. The marginal of x_{kj}^n is given by

$$p(x_{kj}^n|\mathbf{r}) \propto \int_{\mathbf{h},\boldsymbol{\omega},\mathbf{X}\backslash x_{kj}^n} p(\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}|\mathbf{r}), \tag{6.8}$$

where $\mathbf{X}$, $\mathbf{h}$ and $\mathbf{n}$ denote all transmitted symbol, channel taps and colored noise samples. Instead of direct marginalisation, here the joint distribution $p(\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}|\mathbf{r})$ is factorised and factor graph approach is employed to solve the problem in a low complexity way.

According to Bayesian theorem, $p(\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}|\mathbf{r})$ is factorised as

$$p(\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}|\mathbf{r}) \propto p(\mathbf{X})p(\mathbf{h})p(\boldsymbol{\omega})p(\mathbf{r}|\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}). \tag{6.9}$$

Since the transmitted symbols and channel coefficients are independent from each other, $p(\mathbf{X})p(\mathbf{h})$ reads

$$p(\mathbf{X})p(\mathbf{h}) = \prod_{k,j}\left[p(h_{kj})\prod_{n} p(x_{kj}^n)\right], \tag{6.10}$$

where $p(x_{kj}^n)$ is obtained from the output LLR of channel decoder. The prior distribution $p(\boldsymbol{\omega})$ can be factorised based on the AR model as

$$p(\boldsymbol{\omega}) \propto \prod_{j}\prod_{n} \underbrace{\exp(-\frac{\omega_j^n - \sum_{p=1}^{P} a_p \omega_j^{n-p}}{2\sigma_\delta^2})}_{\psi_j^n}, \tag{6.11}$$

Conditioned on ω_j^n, the observation r_j^n at different time n are independent. As shown in [132], using auxiliary variable can help to reduce the computation load. Therefore $p(\mathbf{r}|\mathbf{X}, \mathbf{h}, \boldsymbol{\omega})$ is factorised as

$$p(\mathbf{r}|\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}) \propto \tag{6.12}$$
$$\prod_{j,n} \underbrace{\delta(r_j^n - \sum_{k=1}^{K}\left[h_{kj}\tilde{s}_{kj}^n\right] - \omega_j^n)}_{f_j^n} \cdot \underbrace{\delta(\tilde{s}_{kj}^n - \sum_{i=-L}^{L} g_i x_{kj}^{n-i})}_{\phi_{kj}^n}.$$

Based on the factorisation (6.10)-(6.12), the joint distribution $p(\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}|\mathbf{r})$ can be represented by a factor graph, as shown in Fig. 6.3, on which message passing algorithm is executed to determine the unknown variables.

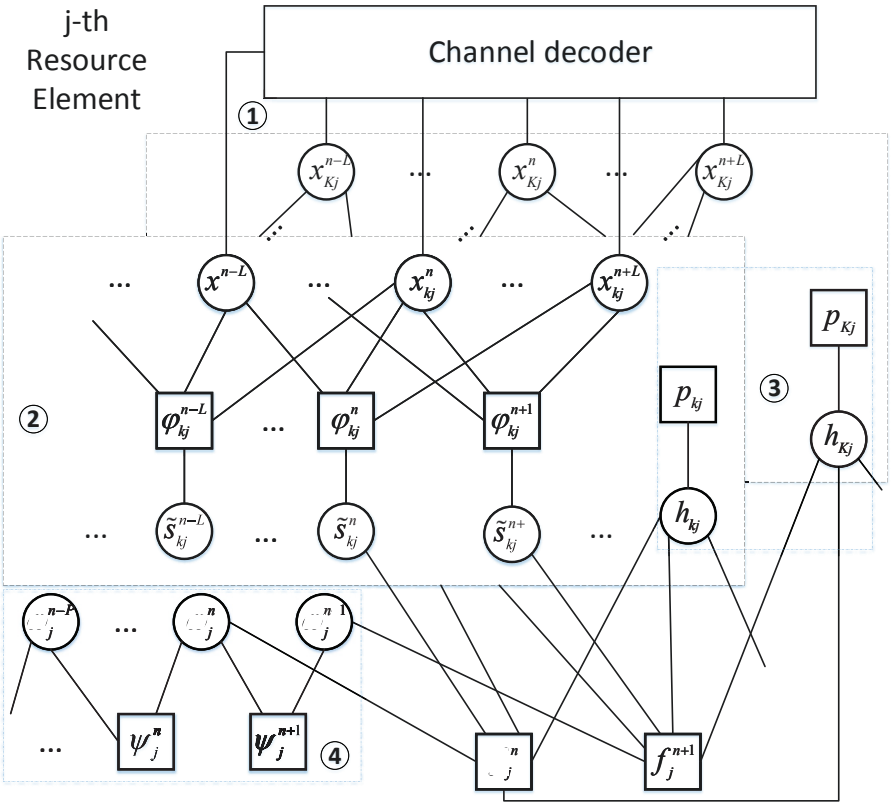


Figure 6.3 : Factor graph representation of the of the jth resource element, where the shorthand notations $p_{kj} = p(h_{kj})$. The factor graph is separated into four parts, i.e. decoding part denoted by ①, equalisation part denoted by ②, channel estimation part denoted by ③, colored noise part denoted by ④.

6.3.3 Message Passing Receiver Design

In this section, the derivations of messages on factor graph of Fig. 6.3 are considered. In the decoding part, the channel decoder and the equalizer exchange extrinsic information iteratively. In the channel decoder, optimal BCJR decoding [47] is utilised. After decoding, the output LLR is

$$L^a(c_{n,m}) = \frac{p(c_{n,m}=0)}{p(c_{n,m}=1)}, \tag{6.13}$$

where the subscripts n and m denotes the nth coded bit and the mth constellation point, respectively. Then the LLRs are transformed to the prior distribution of $p(x_{kj}^n) = \sum_{i=1}^{M} p_i \delta(x_{kj}^n - \chi_i)$, where χ_i is the constellation point in SCMA encoder and p_i is the associated probability. Although the discrete distribution $p(x_{kj}^n)$ can be used as the incoming message in MPA receiver, the complexity will increases exponentially with the number of interfered symbols. Here a Kullback-Leibler divergence based method, also known as expectation propagation (EP) is utilised to approximate the incoming message by Gaussian. EP aims for finding a Gaussian distribution that minimises the Kullback-Leibler divergence, i.e.

$$b_G(x_{kj}^n) = \arg\min_{b_G} \int b_G(x_{kj}^n) \ln \frac{b_G(x_{kj}^n)}{b(x_{kj}^n)} \mathrm{d}x_{kj}^n, \tag{6.14}$$

where b_G belongs to the family of Gaussian distributions and $b(x_{kj}^n)$ is the marginal of variable x_{kj}^n. The minimisation in (6.14) is equivalent to match the moments of $b(x_{kj}^n)$. Assuming the outgoing message has mean and variance $m_{x_{kj}^n}^e$ and $v_{x_{kj}^n}^e$ it is easy to obtain the mean and variance of $b_G(x_{kj}^n)$ as $m_{x_{kj}^n}$ and $v_{x_{kj}^n}$. Then the Gaussian approximation of $p(x_{kj}^n)$ is determined with the mean and variance

$$v_{x_{kj}^n}^0 = \left(\frac{1}{v_{x_{kj}^n}} - \frac{1}{v_{x_{kj}^n}^e} \right)^{-1}, \tag{6.15}$$

$$m_{x_{kj}^n}^0 = v_{x_{kj}^n}^0 \left(\frac{m_{x_{kj}^n}}{v_{x_{kj}^n}} - \frac{m_{x_{kj}^n}^e}{v_{x_{kj}^n}^e} \right). \tag{6.16}$$

Having $m^0_{x^n_{kj}}$ and $v^0_{x^n_{kj}}$, the message in the equalisation part can be calculated. Again, assuming that the message $\mu_{\tilde{s}^n_{kj}\to\phi^n_{k,j}} = \mu_{f^n_j\to\tilde{s}^n_{kj}}$ has been obtained as

$$\mu_{\tilde{s}^n_{kj}\to\phi^n_{k,j}} = \mathcal{G}(m_{\tilde{s}^n_{kj}\to\phi^n_{k,j}}, v_{\tilde{s}^n_{kj}\to\phi^n_{k,j}}), \tag{6.17}$$

then the message $\mu_{\phi^n_{k,j}\to x^{n+l}_{kj}}$ can be written in Gaussian with

$$m_{\phi^n_{k,j}\to x^{n+l}_{kj}} = m_{\tilde{s}^n_{kj}\to\phi^n_{k,j}} - \sum_{i=-L, i\neq l}^{L} g_i m_{x^{n+i}_{kj}\to\phi^n_{k,j}}, \tag{6.18}$$

$$v_{\phi^n_{k,j}\to x^{n+l}_{kj}} = v_{\tilde{s}^n_{kj}\to\phi^n_{k,j}} + \sum_{i=-L, i\neq l}^{L} g_i^2 v_{x^{n+i}_{kj}\to\phi^n_{k,j}}. \tag{6.19}$$

Usually, calculating $\mu_{x^n_{kj}\to\phi^n_{k,j}}$ to different factor nodes $\phi^{n+l}_{k,j}|^L_{l=-L}$ requires to calculate the product of messages for $2L+1$ times. Motivated by the fact that $\mu_{x^n_{kj}\to\phi^n_{k,j}} \cdot \mu_{\phi^n_{k,j}\to x^n_{kj}} = b_G(x^n_{kj})$, the objective message can be calculated in a linear complexity way as $\mu_{x^n_{kj}\to\phi^n_{k,j}} = b_G(x^n_{kj})/\mu_{\phi^n_{k,j}\to x^n_{kj}}$ with

$$v_{x^n_{kj}\to\phi^n_{k,j}} = \left(\frac{1}{v_{x^n_{kj}}} - \frac{1}{v_{\phi^n_{k,j}\to x^n_{kj}}}\right)^{-1}, \tag{6.20}$$

$$m_{x^n_{kj}\to\phi^n_{k,j}} = v_{x^n_{kj}\to\phi^n_{k,j}} \left(\frac{m_{x^n_{kj}}}{v_{x^n_{kj}}} - \frac{m_{\phi^n_{k,j}\to x^n_{kj}}}{v_{\phi^n_{k,j}\to x^n_{kj}}}\right). \tag{6.21}$$

After obtaining all messages $\mu_{\phi^{n+l}_{k,j}\to x^n_{kj}}|^L_{l=-L}$, the mean and variance of the extrinsic message to the channel decoder are given by

$$v^e_{x^n_{kj}} = \left(\sum_{l=-L}^{L} 1/v_{\phi^{n+l}_{k,j}\to x^n_{kj}}\right)^{-1}, \tag{6.22}$$

$$m^e_{x^n_{kj}} = v^e_{x^n_{kj}} \left(\sum_{l=-L}^{L} \frac{m_{\phi^{n+l}_{k,j}\to x^n_{kj}}}{v_{\phi^{n+l}_{k,j}\to x^n_{kj}}}\right). \tag{6.23}$$

Based on $m^e_{x^n_{kj}}$ and $v^e_{x^n_{kj}}$, the extrinsic LLRs are calculated and fed to the channel decoder to determine the data bits of users $\hat{\mathbf{b}}_k$.

Next, let us consider the message updatings in the colored noise part. Since the non-orthogonality of FTN signaling does not affect the first order moment of

noise samples, the mean of noise sample $\mathbb{E}[\omega_j^n] = 0$ holds. According to (6.11), the variance $v_{\psi_j^n \to \omega_j^n}$ is expressed as

$$v_{\psi_j^n \to \omega_j^n} = \sigma_\delta^2 + \sum_{p=1}^{P} (a^p)^2 v_{\omega_j^{n-p} \to \psi_j^n}. \tag{6.24}$$

It should be noted that the colored noise is a causal system that the sample at time n only depends previous noise samples. Therefore the message from ω_j^n to f_j^n is identical with $\mu_{\psi_j^n \to \omega_j^n}$, i.e. $v_{\omega_j^n to f_j^n} = v_{\psi_j^n \to \omega_j^n}$.

For the channel estimation part, the message $\mu_{h_{kj} \to f_j^n}$ is readily determined according to SPA rules as

$$\mu_{h_{kj} \to f_j^n} = p(h_{kj}) \prod_{n' \neq n} \mu_{f_j^{n'} \to h_{kj}}. \tag{6.25}$$

$p(h_{kj})$ is usually coarsely evaluated by using a sequence of pilot symbols, which can be modeled as a Gaussian distributed variable with mean $m_{h_{kj}}^0$ and variance $v_{h_{kj}}^0$. Provided that $\mu_{f_j^{n'} \to h_{kj}}$ has also been obtained in Gaussian form as $\mu_{f_j^{n'} \to h_{kj}} = (m_{f_j^{n'} \to h_{kj}}, v_{f_j^{n'} \to h_{kj}})$, $\mu_{h_{kj} \to f_j^n}$ has mean and variance as

$$m_{h_{kj} \to f_j^n} = v_{h_{kj} \to f_j^n} \left(\frac{m_{h_{kj}}^0}{v_{h_{kj}}^0} + \sum_{n' \neq n} \frac{m_{f_j^{n'} \to h_{kj}}}{v_{f_j^{n'} \to h_{kj}}} \right), \tag{6.26}$$

$$v_{h_{kj} \to f_j^n} = \left(\frac{1}{v_{h_{kj}}^0} + \sum_{n' \neq n} \frac{1}{v_{f_j^{n'} \to h_{kj}}} \right)^{-1}. \tag{6.27}$$

The belief $b(h_{kj})$ is obtained by adding the terms with index $n^{'} = n$ into (6.26) and (6.27). And the maximum *a posteriori* (MAP)estimator can be user to determine the estimate of channel coefficient by $\hat{h}_{kj} = \arg\max_{h_{kj}} b(h_{kj})$. Since $b(h_{kj})$ is Gaussian distribution, the MAP estimate $\hat{h}_{kj}$ is the mean of $b(h_{kj})$

In the above, messages have been derived in Gaussian closed form in four parts of the factor graph. However, they are based on the fact that the messages from f_j^n to its connected variable vertices are Gaussian distributions. In what follows, the

messages related to vertex f_j^n will be computed. The message $\mu_{f_j^n \to \tilde{s}_{kj}^n}$ is expressed as

$$\begin{aligned}\mu_{f_j^n \to \tilde{s}_{kj}^n} &\propto \int \delta(r_j^n - \sum_{k=1}^{K}[h_{kj}\tilde{s}_{kj}^n] - \omega_j^n)\mu_{\omega_j^n \to f_j^n} \prod_k \mu_{h_{kj} \to f_j^n} \prod_{k' \neq k} \mu_{\tilde{s}_{k'j}^n \to f_j^n} \mathrm{d}h_{kj} \mathrm{d}\omega_j^n \mathrm{d}\tilde{s}_{k'j}^n \\ &\propto \int \exp\left(-\frac{|r_j^n - \sum_{k=1}^{K}[h_{kj}\tilde{s}_{kj}^n]|^2}{v_{\omega_j^n \to f_j^n}}\right) \prod_k \mu_{h_{kj} \to f_j^n} \prod_{k' \neq k} \mu_{\tilde{s}_{k'j}^n \to f_j^n} \mathrm{d}h_{kj} \mathrm{d}\tilde{s}_{k'j}^n \\ &\propto \int \exp\left(-\frac{|r_j^n - \sum_{k=1}^{K}[m_{h_{kj} \to f_j^n}\tilde{s}_{kj}^n]|^2}{v_{\omega_j^n \to f_j^n} + \sum_{k=1}^{k}|\tilde{s}_{kj}^n|^2 v_{h_{kj} \to f_j^n}}\right) \prod_{k' \neq k} \mu_{\tilde{s}_{k'j}^n \to f_j^n} \mathrm{d}\tilde{s}_{k'j}^n. \end{aligned} \tag{6.28}$$

From (6.28), it is seen when calculating message $\mu_{f_j^n \to \tilde{s}_{kj}^n}$, the variable $\tilde{s}_{kj}^n$ appears in both numerator and denominator of the exponent term, which makes the conventional MPA unavailable. Here variational message passing (VMP) in [78] is used where the message from factor vertex f to variable vertex x is formulated as

$$\mu_{f \to x}(x) \propto \exp\left(\int \ln f(\mathbf{x}) \prod_{x' \in \mathcal{S}(f) \backslash \{x\}} \mu_{x' \to f}(x') \mathrm{d}x'\right). \tag{6.29}$$

Consequently, the message (6.28) can be obtained in Gaussian form with mean

$$m_{f_j^n \to \tilde{s}_{kj}^n} = \frac{(r_j^n - \sum_{k'=1, k \neq k}^{K} m_{h_{k'j} \to f_j^n} m_{\tilde{s}_{k'j}^n \to f_j^n}) m_{h_{kj} \to f_j^n}}{|m_{h_{kj} \to f_j^n}|^2 + v_{h_{kj} \to f_j^n}}, \tag{6.30}$$

and variance

$$v_{f_j^n \to \tilde{s}_{kj}^n} = \frac{v_{\omega_j^n \to f_j^n}}{|m_{h_{kj} \to f_j^n}|^2 + v_{h_{kj} \to f_j^n}}. \tag{6.31}$$

The message $\mu_{f_j^n \to h_{kj}}$ can be calculated by VMP likewise, whose mean and variance are

$$m_{f_j^n \to h_{kj}} = \frac{(r_j^n - \sum_{k'=1, k \neq k}^{K} m_{h_{k'j} \to f_j^n} m_{\tilde{s}_{k'j}^n \to f_j^n}) m_{\tilde{s}_{kj}^n \to f_j^n}}{|m_{\tilde{s}_{kj}^n \to f_j^n}|^2 + v_{\tilde{s}_{kj}^n \to f_j^n}}, \tag{6.32}$$

$$v_{f_j^n \to h_{kj}} = \frac{v_{\omega_j^n \to f_j^n}}{|m_{\tilde{s}_{kj}^n \to f_j^n}|^2 + v_{\tilde{s}_{kj}^n \to f_j^n}}. \tag{6.33}$$

6.3.4 Algorithm Summary

Using appropriate approximations, all messages on the factor graph are represented in parametric forms, which reduce the computational complexity of the conventional MPA receiver significantly. Compared to some existing advanced MPA

receiver, due to the introduction of auxiliary variable and modified message updating rules, the computational complexity increases only linearly with the number of users and resource elements. The details of the proposed receiver for joint channel estimation and decoding in FTN-SCMA system are organised in Algorithm 6.

Algorithm 6 Joint Channel Estimation and Decoding Algorithm for FTN-SCMA System

Initialisation:

At the first turbo iteration, initialise all undetermined messages as Gaussian distribution with zero mean and unit variance;

Using pilot sequence to coarsely estimate the mean $m^0_{h_{kj}}$ and variance $v^0_{h_{kj}}$ of channel coefficient.

for iter=1:N_{iter} **do**

  Compute the means and variances of messages in equalisation part according to (6.18)-(6.21);

  Compute the message from factor vertex f^n_j to variable vertices x^n_{kj} and h_{kj} according to (6.30)-(6.33);

  Compute the variance $v_{\psi^n_j \to \omega^n_j}$ according to (6.24);

  Compute the message from h_{kj} to factor vertex f^n_j via (6.26) and (6.27);

  Compute the mean and variance of message to channel decoder according to (6.22) and (6.23);

  Convert the outgoing messages to LLR and feed them to the channel decoder;

  Perform BCJR decoding;

  Convert the extrinsic LLRs to Gaussian messages using (6.15) and (6.16);

end for

Determine the estimate of channel coefficient by MAP estimator.

6.4 User Activity Detection in Grant-free FTN-SCMA Systems

In Grant-free system, the user do not need grant before sending signals to the base station. In existing works, a precision parameter is used as a hyper-prior to capture the channel sparsity. However, this makes the factor graph contains more short loops and increase the receiver complexity. In this section, an algorithm for FTN-SCMA system that determines the user activity directly while perform channel estimation and decoding will be proposed.

Let us use a binary variable $\xi_k = \{0, 1\}$ to denote the user activity, i.e. $\xi_k = 1$ indicates the user k is active and vice versa. Then the nth sample at the jth resource element is expressed by

$$r_j^n = \sum_{k=1}^{K} h_{kj}\xi_k\tilde{s}_{kj}^n + \omega_j^n. \tag{6.34}$$

The prior distribution of ξ_k is a Bernoulli distribution given by

$$p^0(\xi_k) = p_1^{\xi_k}(1-p_1)^{1-\xi_k}, \tag{6.35}$$

where p_1 is the prior knowledge of user activity based on existing data.

6.4.1 Probability based Active User Detection Algorithm

To determine the activity of user k, the probability of it being active γ_k should be calculated based on the receiver samples. To this end, the factor graph structure 6.4 is modified by including ξ_k. Since only function f is affected, only the corresponding parts of the factor graph are illustrated, as shown in Fig. 6.4. Here an auxiliary variable ξ_{kj} is used on the edge connected vertices h_{kj}.[2] Then it is possible to formulate the message passing algorithm to calculate the probability of $\xi_{kj} = 1$, namely γ_{kj}.

[2]Alternatively, the variable ξ can be put on the edge connected $\tilde{s}_{kj}^n$ and f_j^n or the edge connected h_{kj} and f_j^n. However, these two means will increases N times variables.

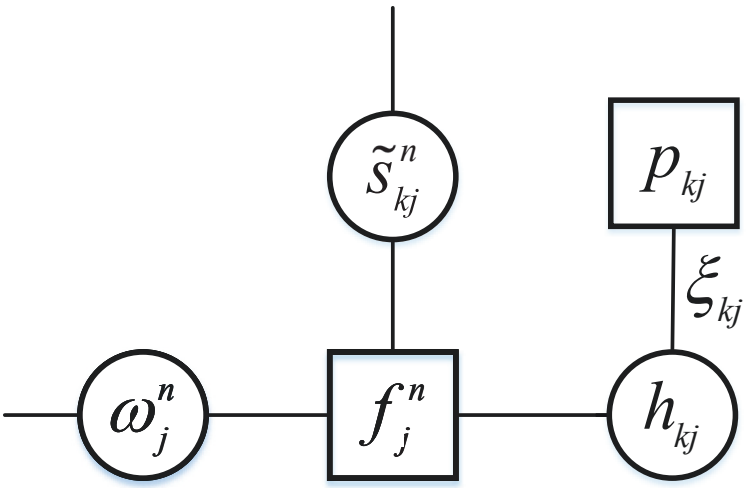


Figure 6.4 : Modified factor graph structure including user activity.

Based on Fig. 6.4, the message from f_j^n to h_{kj} is obtained with mean $m_{f_j^n \to h_{kj}}$ and variance $v_{f_j^n \to h_{kj}}$ according to (6.32) and (6.33). Hence the intrinsic message for $\xi_{kj}h_{kj}$ is obtained with mean and variance

$$\overrightarrow{m}_{\xi_{kj}h_{kj}} = \overrightarrow{v}_{\xi_{kj}h_{kj}} \sum_n \frac{m_{f_j^n \to h_{kj}}}{v_{f_j^n \to h_{kj}}}, \tag{6.36}$$

$$\overrightarrow{v}_{\xi_{kj}h_{kj}} = \left(\sum_n \frac{1}{v_{f_j^n \to h_{kj}}} \right)^{-1}. \tag{6.37}$$

The distribution of ξ_{kj} is obtained by integrating h_{kj} over the joint distribution, formulating as

$$\begin{aligned} p(\xi_{kj}) &\propto \int \exp\left(-\frac{(\xi_{kj}h_{kj} - \overrightarrow{m}_{\xi_{kj}h_{kj}})^2}{\overrightarrow{v}_{\xi_{kj}h_{kj}}} \right) \exp\left(-\frac{(h_{kj} - m_{h_{kj}}^0)^2}{v_{h_{kj}}^0} \right) \mathrm{d}h_{kj} \\ &\propto \exp\left(-\frac{(\xi_{kj}m_{h_{kj}}^0 - \overrightarrow{m}_{\xi_{kj}h_{kj}})^2}{\xi_{kj}^2 v_{h_{kj}}^0 + \overrightarrow{v}_{\xi_{kj}h_{kj}}} \right). \end{aligned} \tag{6.38}$$

Then the probability γ_{kj} is updated as

$$\begin{aligned} \gamma_{kj} &= \frac{p(\xi_{kj}=1)}{p(\xi_{kj}=0) + p(\xi_{kj}=1)} \\ &= \frac{1}{1 + \frac{p(\xi_{kj}=0)}{p(\xi_{kj}=1)}}. \end{aligned} \tag{6.39}$$

After obtaining the probability of $\xi_{kj} = 1$, it is readily to obtain the probability γ_k as

$$\gamma_k = \frac{\prod_j \gamma_{kj} p_1}{\prod_j \gamma_{kj} p_1 + \prod_j (1 - \gamma_{kj})(1 - p_1)}. \tag{6.40}$$

To determine the value of ξ_k, a threshold β is set according to empirical evidence. Then we say the user k is active if $\gamma_k \geq \beta$ and vice versa.

The extrinsic message from h_{kj} to f_j^n is still obtained by $\mu_{h_{kj} \to f_j^n} = \mu_{p_{kj} \to h_{kj}} \prod_{n' \neq n} \mu_{f_j^{n'} \to h_{kj}}$. Specifically when calculating $\mu_{p_{kj} \to h_{kj}}$, ξ_{kj} and h_{kj} are combined together as a new variable,

$$\mu_{p_{kj} \to h_{kj}} \propto \left(\gamma_{kj} e^{-\frac{(h_{kj} - m^0_{h_{kj}})^2}{v^0_{h_{kj}}}} + (1 - \gamma_{kj}) e^{-\frac{[m^0_{h_{kj}}]^2}{v^0_{h_{kj}}}} \right). \tag{6.41}$$

Obviously, $\mu_{p_{kj} \to h_{kj}}$ is a Gaussian mixture distribution (GMD) and $\mu_{h_{kj} \to f_j^n}$ is aslo a GMD. In conjunction with the message passing receiver in Section III, $\mu_{h_{kj} \to f_j^n}$ is approximated to Gaussian by determining its mean and variance by

$$m_{h_{kj} \to f_j^n} = \mathbb{E}_{\mu_{h_{kj} \to f_j^n}}[h_{kj}], \tag{6.42}$$

$$v_{h_{kj} \to f_j^n} = \mathbb{E}_{\mu_{h_{kj} \to f_j^n}}[h_{kj}^2] - m^2_{h_{kj} \to f_j^n}. \tag{6.43}$$

In the following table, the algorithm for user activity detection based on message passing algorithm is described. To sum up, it is seen that the extension from the proposed algorithm in Section III is readily and only small modification is required. However, since ξ_{kj} has to be calculated separately, the receiver complexity increases. Also, the derivation of messages is not intuitive from the perspective of probabilistic factorisation. In the following subsection, another active user detection method with reduced complexity will be proposed.

6.4.2 Message Passing based Active User Detection Algorithm

To achieve a concise form of message passing receiver under the factor graph framework, ξ_k is added as a new variable vertex on the factor graph. According to

Algorithm 7 User Activity Detection Algorithm I

1: Run Algorithm 6;

2: Calculate the intrinsic message to h_{kj} according to (6.36) and (6.37);

3: Determine the probability γ_{kj} by (6.39);

4: Calculate γ_k according to (6.40) and decide ξ_k;

5: Approximate the message $\mu_{h_{kj}\to f_j^n}$ to Gaussian and continue running algorithm 1.

(6.34), a dirac Delta function $\delta(\bar{s}_{kj}^n - \xi_k \tilde{s}_{kj}^n)$ is used to represent the multiplication relationship of $\bar{s}_{kj}^n = \xi_k \tilde{s}_{kj}^n$. Accordingly, the joint likelihood function (6.12) is revised as

$$p(\mathbf{r}|\mathbf{X},\mathbf{h},\boldsymbol{\omega},\boldsymbol{\xi}) \propto \prod_{j,n} \underbrace{\delta(r_j^n - \sum_{k=1}^{K} \left[h_{kj}\bar{s}_{kj}^n\right] - \omega_j^n)}_{f_j^n} \cdot \delta(\bar{s}_{kj}^n - \xi_k \tilde{s}_{kj}^n) \cdot \underbrace{\delta(\tilde{s}_{kj}^n - \sum_{i=-L}^{L} g_i x_{kj}^{n-i})}_{\phi_{kj}^n}, \tag{6.44}$$

and the factor graph is modified as shown in Fig. 6.5. Since ξ_k is a binary variable, which follows a discrete distribution. Then in message updating, when the messages from different product verteices to ξ_k are Gaussian, the message $\mu_{\xi_k\to\times_{kj}^n}$ follows a Gaussian mixture distribution, which makes it unavailable to derive Gaussian messages. To tackle this problem, the message from ξ_k to the product vertex $\times_{kj}^n$ is approximated by Gaussian via expectation propagation.

Following SPA rules, the belief of ξ_k is $b(\xi_k) = \mu_{\xi_k\to p_{\xi_k}} p(\xi_k)$. Assuming $\mu_{\xi_k\to p_{\xi_k}}$ is Gaussian with mean $m_{\xi_k\to p_{\xi_k}}$ and variance $v_{\xi_k\to p_{\xi_k}}$, the mean and variance of $b(\xi_k)$

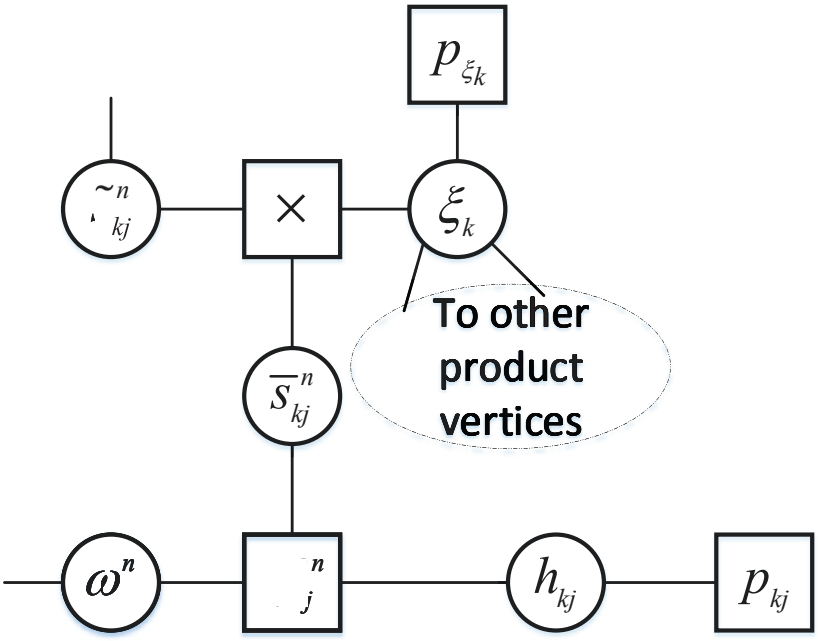


Figure 6.5 : Modified factor graph structure. The product node $\times^n_{kj}$ represents the constraint $\delta(\bar{s}^n_{kj} - \xi_k \tilde{s}^n_{kj})$.

is

$$m_{\xi_k} = \frac{p_1 \exp(-\frac{1-2m_{\xi_k \to p_{\xi_k}}}{v_{\xi_k \to p_{\xi_k}}})}{p_1 \left[\exp(-\frac{1-2m_{\xi_k \to p_{\xi_k}}}{v_{\xi_k \to p_{\xi_k}}}) - 1\right] + 1}, \tag{6.45}$$

$$v_{\xi_k} = m_{\xi_k} - m^2_{\xi_k}. \tag{6.46}$$

An obvious observation is that in (6.45) the absolute value of the exponent term dominates the value of m_{ξ_k}, and v_{ξ_k} becomes smaller when ξ_k is approaching 0 or 1. That is to say after running several iterations, the belief of ξ_k becomes more 'concentrated'. Having m_{ξ_k} and v_{ξ_k}, the Gaussian approximation of message $\mu_{\xi_k \to \times^n_{kj}}$ is easily determined as

$$\mu_{\xi_k \to \times^n_{kj}} \sim \mathcal{G}\left(\frac{m_{\xi_k} v_{\xi_k} - m_{\times^n_{kj} \to \xi_k} v_{\times^n_{kj} \to \xi_k}}{v_{\xi_k} - v_{\times^n_{kj} \to \xi_k}}, \frac{v_{\xi_k} v_{\times^n_{kj} \to \xi_k}}{v_{\xi_k} - v_{\times^n_{kj} \to \xi_k}}\right). \tag{6.47}$$

For the product vertex, as $\mu_{\xi_k \to \times^n_{kj}}$ and $\mu_{\tilde{s}^n_{kj} \to \times^n_{kj}} = \mu_{\phi^n_{kj} \to \tilde{s}^n_{kj}}$ are available , the

mean and variance of message $\mu_{\bar{s}_{kj}^n \to f_j^n}$ are given as

$$m_{\bar{s}_{kj}^n \to f_j^n} = m_{\xi_k \to \times_{kj}^n} m_{\phi_{kj}^n \to \tilde{s}_{kj}^n}, \tag{6.48}$$

$$m_{\bar{s}_{kj}^n \to f_j^n} = v_{\xi_k \to \times_{kj}^n} m_{\phi_{kj}^n \to \tilde{s}_{kj}^n}^2 + (m_{\xi_k \to \times_{kj}^n}^2 + v_{\xi_k \to \times_{kj}^n}) v_{\phi_{kj}^n \to \tilde{s}_{kj}^n}. \tag{6.49}$$

The detailed derivations of (6.48) and (6.49) are given in the Appendix F. The message $\mu_{\bar{s}_{kj}^n \to \times_{kj}^n}$, conversely, is same as the message calculated in (6.30) and (6.31). Next the messages from $\times_{kj}^n$ to $\tilde{s}_{kj}^n$ and ξ_k are calculated. Again, a similar problem as in section occurs that even $\mu_{\bar{s}_{kj}^n \to \times_{kj}^n}$ is Gaussian, it is not possible to formulate Gaussian form messages for ξ_k. To overcome this challenge, the message $\mu_{m_{\bar{s}_{kj}^n} \to \times_{kj}^n}$ and the constraint are grouped as a new soft node. Then, the joint distribution $p(m_{\tilde{s}_{kj}^n}, \xi_k)$ is formulated as

$$p(m_{\tilde{s}_{kj}^n}, \xi_k) \propto \exp\left(-\frac{(m_{\bar{s}_{kj}^n \to \times_{kj}^n} - \xi_k \tilde{s}_{kj}^n)^2}{v_{\bar{s}_{kj}^n \to \times_{kj}^n}} \right) \mu_{\xi_k \to \times_{kj}^n} \mu_{\tilde{s}_{kj}^n \to \times_{kj}^n}. \tag{6.50}$$

According the variational inference framework, $b(\xi_k)b(\tilde{s}_{kj}^n)$ is used to approximate (6.50). The Kullback Leibler divergence is given as

$$\begin{aligned} \mathrm{KLD}(\xi_k, \tilde{s}_{kj}^n) &= \int b(\xi_k) b(\tilde{s}_{kj}^n) \ln \frac{b(\xi_k) b(\tilde{s}_{kj}^n)}{p(\tilde{s}_{kj}^n, \xi_k)} \mathrm{d}\xi_k \mathrm{d}\tilde{s}_{kj}^n \\ &= -\int b(\xi_k) \left[\int \ln p(\tilde{s}_{kj}^n, \xi_k) b(\tilde{s}_{kj}^n) \mathrm{d}\tilde{s}_{kj}^n \right] \mathrm{d}\xi_k \\ &\quad + \int b(\xi_k) \ln b(\xi_k) \mathrm{d}\xi_k + C, \end{aligned} \tag{6.51}$$

where C denotes a constant. To minimise the KLD, it is easily seen

$$b(\xi_k) = \exp\left(\int \ln p(\tilde{s}_{kj}^n, \xi_k) b(\tilde{s}_{kj}^n) \right). \tag{6.52}$$

Substituting (6.50) into (6.52) yields

$$\frac{b(\xi_k)}{\mu_{\xi_k \to \times_{kj}^n}} \propto \exp\left(-\xi_k^2 \frac{m_{\tilde{s}_{kj}^n}^2 + v_{\tilde{s}_{kj}^n}}{v_{\bar{s}_{kj}^n \to \times_{kj}^n}} + 2\xi_k \frac{m_{\bar{s}_{kj}^n \to \times_{kj}^n} m_{\tilde{s}_{kj}^n}}{v_{\bar{s}_{kj}^n \to \times_{kj}^n}} \right), \tag{6.53}$$

where $v_{\tilde{s}_{kj}^n} = (v_{\tilde{s}_{kj}^n \to \times_{kj}^n}^{-1} + v_{\times_{kj}^n \to \tilde{s}_{kj}^n}^{-1})^{-1}$ and $m_{\tilde{s}_{kj}^n} = v_{\tilde{s}_{kj}^n}(m_{\tilde{s}_{kj}^n \to \times_{kj}^n} v_{\tilde{s}_{kj}^n \to \times_{kj}^n}^{-1} + m_{\times_{kj}^n \to \tilde{s}_{kj}^n} v_{\times_{kj}^n \to \tilde{s}_{kj}^n}^{-1})$.

Therefore the message $\mu_{\times_{kj}^n \to \xi_k}$ is determined in Gaussian with mean and variance

$$m_{\times_{kj}^n \to \xi_k} = \frac{m_{\bar{s}_{kj}^n \to \times_{kj}^n} m_{\tilde{s}_{kj}^n}}{m_{\tilde{s}_{kj}^n}^2 + v_{\tilde{s}_{kj}^n}}, \tag{6.54}$$

$$v_{\times_{kj}^n \to \xi_k} = \frac{v_{\bar{s}_{kj}^n \to \times_{kj}^n}}{m_{\tilde{s}_{kj}^n}^2 + v_{\tilde{s}_{kj}^n}}. \tag{6.55}$$

Similarly, the Gaussian message $\mu_{\times kj^n \to \tilde{s}_{kj}^n}$ is obtained as

$$\mu_{\times kj^n \to \tilde{s}_{kj}^n} \propto \mathcal{G}\left(\frac{m_{\xi_k \to \times_{kj}^n} m_{\xi_k}}{m_{\xi_k}^2 + v_{\xi_k}}, \frac{v_{\xi_k \to \times_{kj}^n}}{m_{\xi_k}^2 + v_{\xi_k}} \right), \tag{6.56}$$

where m_{ξ_k} and v_{ξ_k} are the mean and variance of $b(\xi_k)$. Having $\mu_{\times_{kj}^n \to \xi_k}$ in Gaussian form, the mean and variance of Gaussian message $\mu_{\xi_k \to p_{\xi_k}}$ can be calculated by straightforward manipulations. The value of ξ_k is given by the MAP estimate of $b(\xi_k)$, which is shown in (6.45). Again, a threshold β is set and comparing it with m_{ξ_k} gives the decision whether user k is active or inactive. The detailed of the proposed user activity detection algorithm II is summarised in Algorithm. 8.

Algorithm 8 User Activity Detection Algorithm II

1: Run Algorithm 6;
2: Approximate the message from ξ_k to the product vertex to Gaussian by EP according to (6.45) and (6.47);
3: Calculate the mean $m_{\bar{s}_{kj}^n \to f_j^n}$ and variance $v_{\bar{s}_{kj}^n \to f_j^n}$ using (6.48) and (6.49);
4: Determine the messages from the product vertex to ξ_k and $\tilde{s}_{kj}^n$ using (6.54)-(6.56);
5: Calculate the message $\mu_{\xi_k \to p_{\xi_k}}$ and estimate ξ_k using (6.45);
6: Continue running Algorithm 1.

6.5 Simulation Results

In this section, the performance of the proposed algorithm is evaluated via simulations. An SCMA system with $J = 4$ resource elements that supports $K = 6$

users is considered. The codebook is defined according to [103] with size $M = 4$ and indicator matrix $\mathbf{F}$ as

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}. \tag{6.57}$$

Each user transmits a sequence of data bits, which is coded using a rate-5/7 low density parity code (LDPC) and then mapped to a sequence of SCMA codewords. The number of transmitted symbols corresponding to each user is $N = 2048$. The transmitted symbols pass through root raised cosine shaping filters with roll-off factor $\alpha = 0.5$ and packing factor $\tau = 0.8$.[3] The number of interfered symbols is assumed to be $L = 10$ on both sides. The channel is set to be Rayleigh fading whose impulse response is generated according to Jake's model. The coarse estimate of channel coefficients is obtained by using 5 pilots symbols. The maximum number of iterations between the detector and the channel decoder is $N_{iter} = 10$. All results are averaged from 1000 independent Monte Carlo trails.

In Fig. 6.6, the proposed algorithm is compared with the MPA-Gauss, and MMSE-MPA methods in terms of bit error ratio (BER). As reference, the performance of conventional MPA receiver under maximum *a posteriori* criterion is illustrated. The 'MPA-Gauss' method refers to the method that directly approximates the prior distribution of $p(x_{kj}^n)$ by Gaussian distribution. The 'MMSE-MPA' method is a combination of MMSE equalizer to eliminate the effect of channel fading and ISI and SCMA decoder to decode the SCMA codewords. It is observed

[3]It is assumed the same shaping filter is employed for different resource elements at the transmitter side.

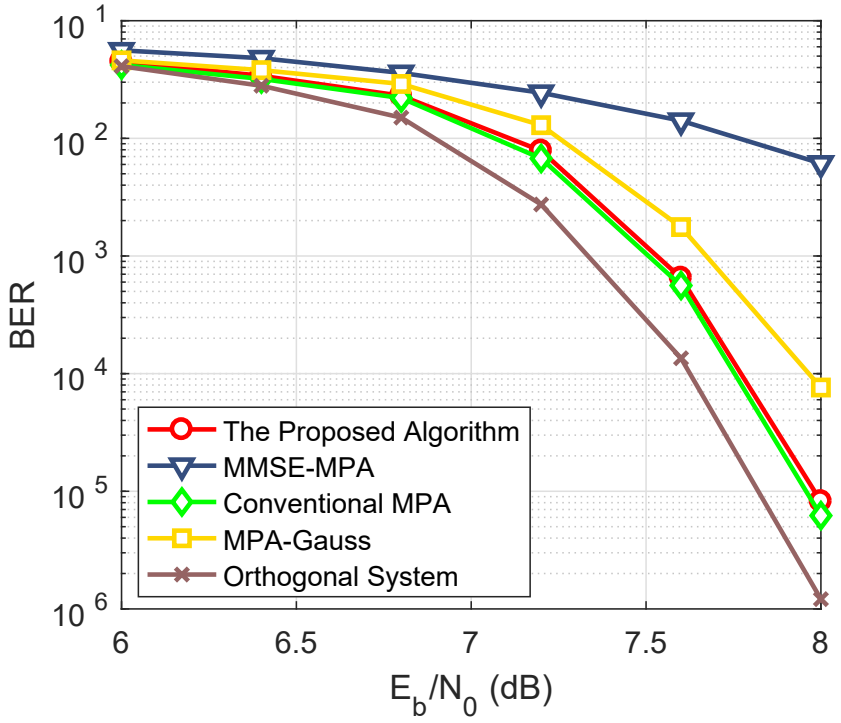


Figure 6.6 : BER performance of different algorithms for FTN-SCMA system.

the proposed algorithm outperforms all three algorithms and has almost the same performance as the conventional MPA receiver. The 'MMSE-MPA' method suffers from significant performance loss due to the error propagation. Moreover, using MMSE equalizer impose a cubic order of complexity, which is prohibitively high in practical applications. Compared to 'MPA-Gauss' method, since EP further use the extrinsic information fed to the channel decoder, the proposed algorithm achieves performance gain. Moreover, the performance based on an OMA system using Nyquist signaling is plotted. It is observed that the performance loss of the proposed algorithm is as small as 0.2dB. Meanwhile, 50%more user are supported and 25% higher data rate is achieved. That is to say, using the same resources, total 87.5% more information can be transmitted via the considered FTN-SCMA system with negligible performance loss.

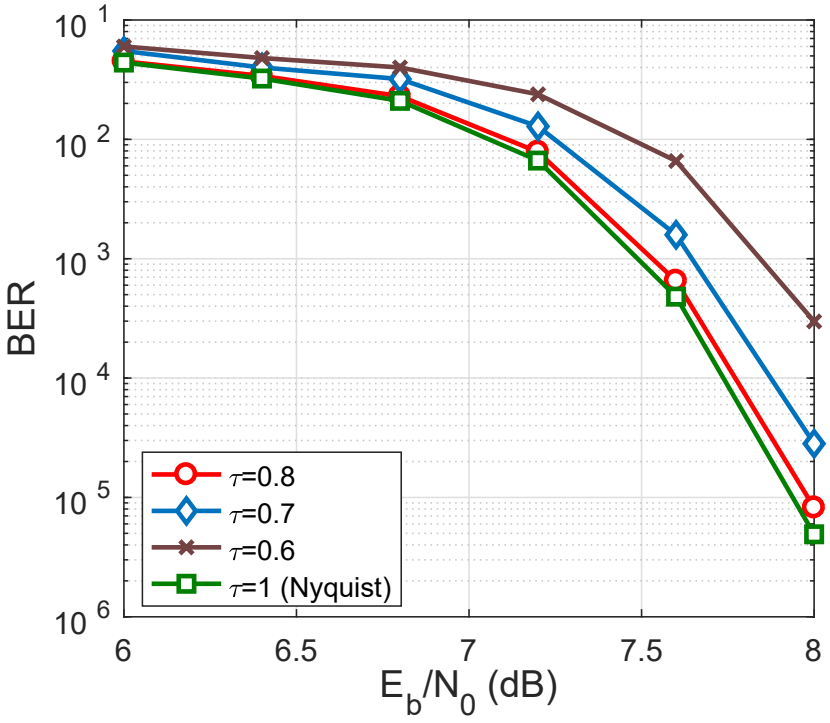


Figure 6.7 : BER performance of the proposed algorithm with different τ.

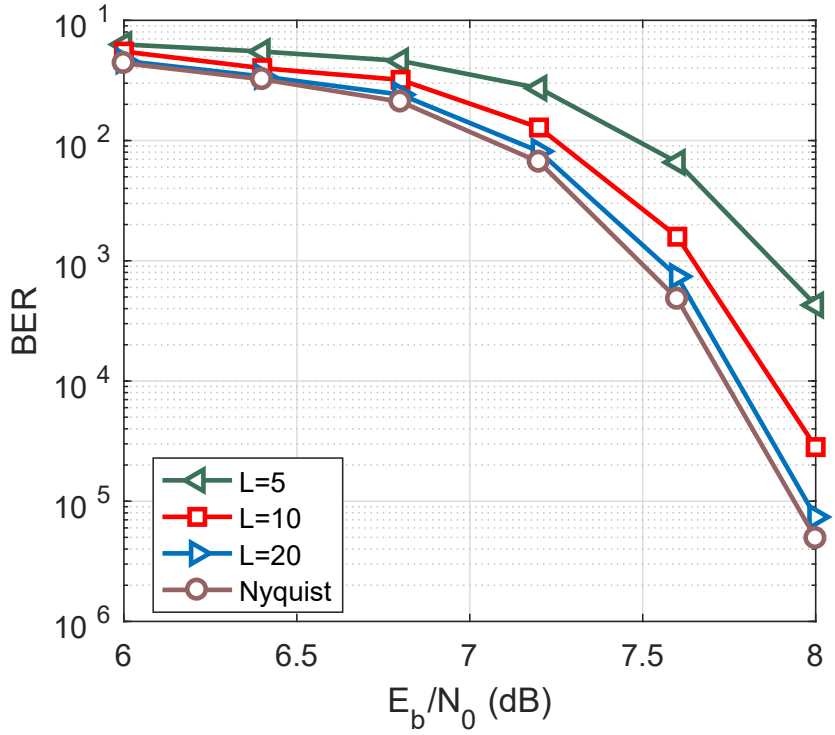


Figure 6.8 : BER performance of the proposed algorithm with different L.

Fig. 6.7 depicts the BER versus E_b/N_0 of the proposed algorithm parameterised by different packing factor τ, where $\tau = 1$ corresponds to the Nyquist signaling case. It is seen that the proposed iterative receiver for FTN-SCMA system is capable of achieving similar performance of the Nyquist signaling case when $\tau \geq 0.8$. Moreover, as the packing ratio becomes smaller, severer interference is introduced and the performance gap between the FTN signaling and Nyquist signaling becomes larger. Actually, another reason maybe in (6.5) the number of interfered symbols L is not enough to describe the ISI induced by FTN. In Fig. 6.8, the BER curves with various values of L are illustrated while the packing ratio τ is fixed as 0.7. It is observed when L increases to 20, the performance gap between the FTN signaling case and Nyquist one becomes negligible, which means using a smaller packing ratio is still possible at the cost of more complex receiver. This implies the transmission rate and the receiver complexity can be compromised. Nevertheless, due to the existence of Mazo limit, the packing factor can not be decreased boundlessly.

The BER performance of the proposed algorithm versus the number of iterations is illustrated in Fig. 6.9 to see the convergence behavior. It can be seen with different values of E_b/N_0, the proposed algorithm can converge after several iterations. Moreover, it is noted that with higher value of E_b/N_0, the proposed algorithm requires more iterations to guarantee the convergence. In Fig. 6.10, the normalised mean squared error (NMSE) of the estimated channel coefficients versus E_b/N_0 is illustrated. The NMSE is defined as

$$\text{NMSE}_h = \frac{\sum_{k=1}^{K} \|\mathbf{h}_k - \hat{\mathbf{h}}_k\|^2}{\sum_{k=1}^{K} \|\mathbf{h}_k\|^2}, \tag{6.58}$$

where $\hat{\mathbf{h}}_k$ is the channel estimate obtained in Section III. The NMSEs of the least square channel estimation method using 5 and all pilot symbols are depicted for reference. From Fig. 6.10, it is seen that the proposed algorithm is efficient in channel estimation, which can attain the performance of the LS algorithm based

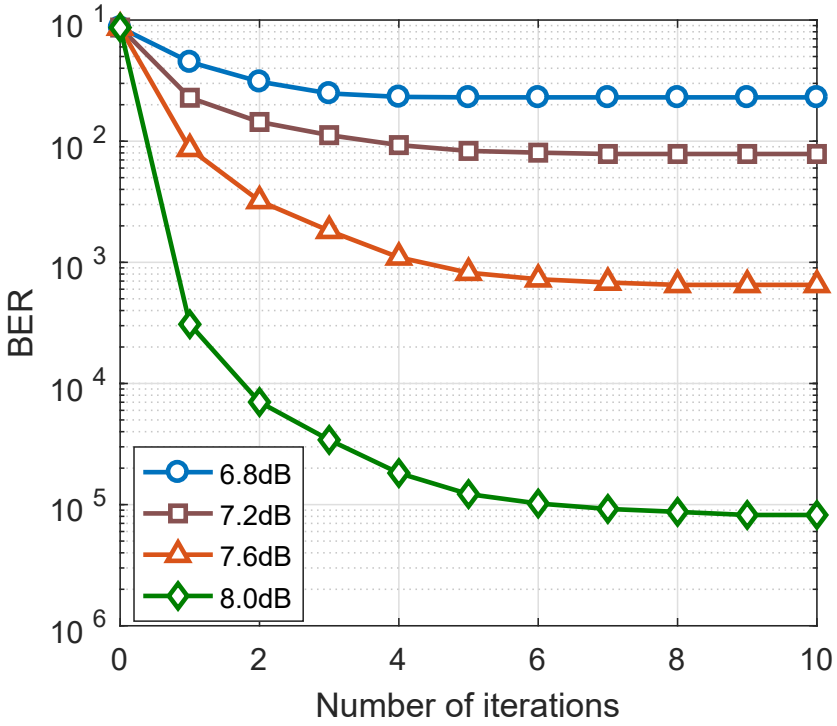


Figure 6.9 : BER performance of the proposed algorithm versus the number of iterations.

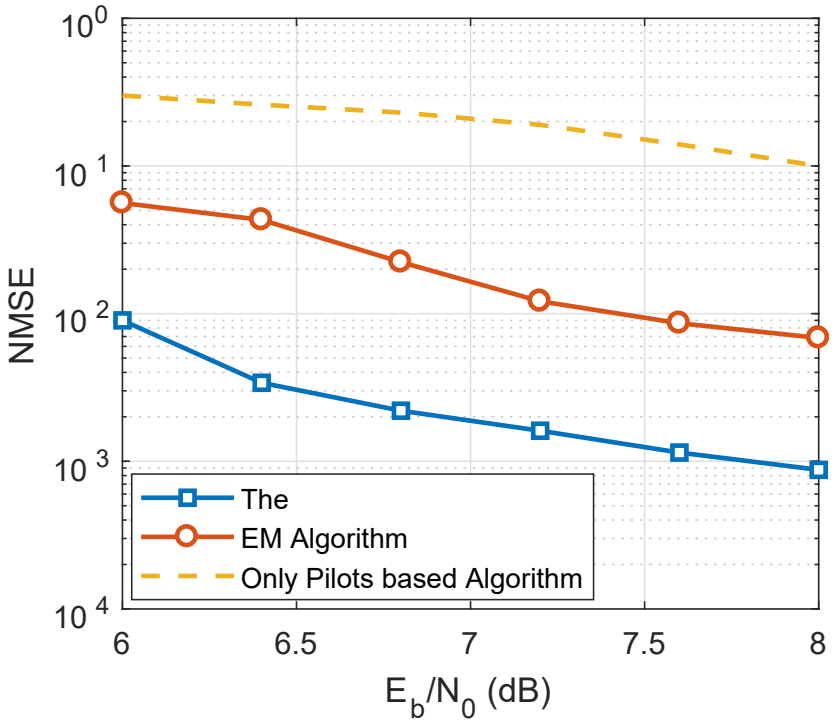


Figure 6.10 : NMSEs of different algorithms versus E_b/N_0.

on all pilot symbols. Compared to the coarse estimate with only 5 pilot symbols, the proposed algorithm has significantly improved the channel estimation result. Moreover, the performance of an advanced joint channel estimation and decoding algorithm based on expectation maximisation (EM) is included here. Since EM will discard uncertainties of variables in the iterative process, it suffers from performance loss.

Next, the performance of the proposed two active user detection algorithms is evaluated in a grant-free system. In Fig. 6.11, the BER performance of the proposed algorithm versus E_b/N_0 is illustrated, where the probability that user is active is $p_1 = 0.3$. For comparison, the performances of the proposed algorithm in Section III with known active users (denoted by 'MPA-Known'), the algorithm that regards all users as active users (denoted by 'Approx-known') and the two-stage CS-MPA

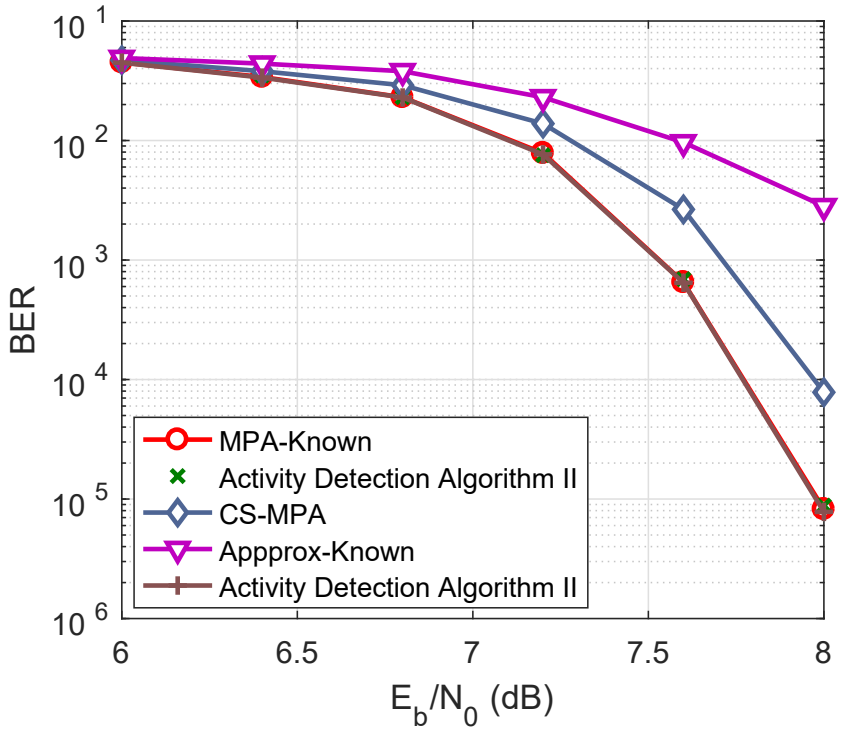


Figure 6.11 : BER performance of the proposed active user detection algorithms and existing method.

algorithm [133] that uses compressive sensing for active user detection first and then performs MPA multiuser detector. It can be observed that 'Approx-known' that assumes all users to be active suffers from significant performance degradation. Due to that the two stage method only provides hard decision of active users to the equalisation part, it also experiences considerable performance loss. Compared to the optimal case that all users' activties are known, the proposed algorithms designed under factor graph framework can achieve near optimal performance. Since the user activity detection algorithm II has lower complexity than algorithm I, it is more attractive in practical grant-free systems.

Fig. 6.12 depicts the NMSE of channel estimate based on the joint channel estimation, decoding and active user detection algorithm parameterised by the oc-

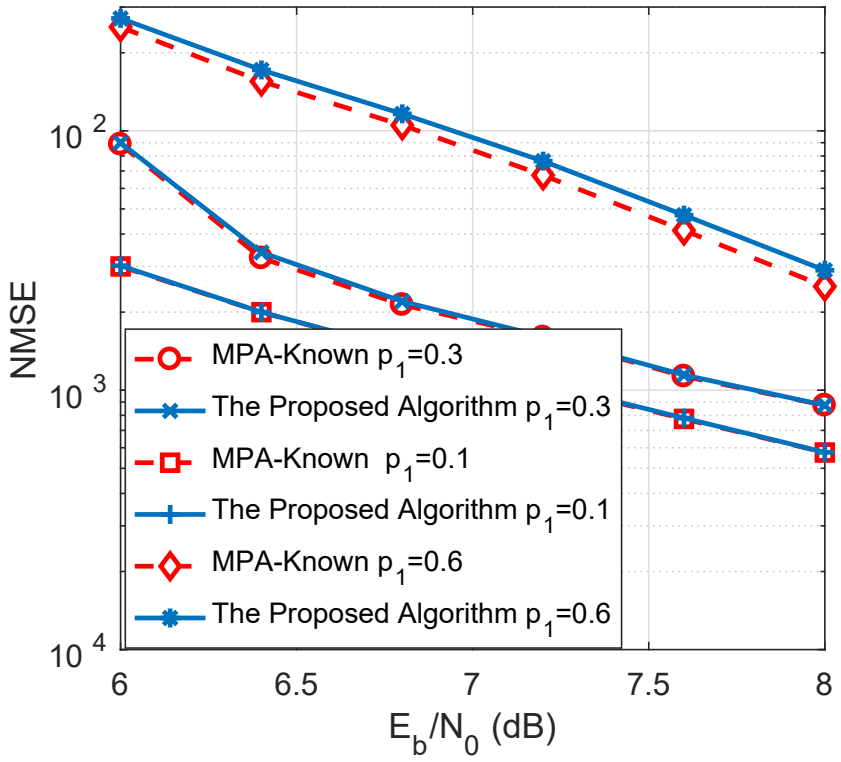


Figure 6.12 : NMSE of channel estimate with different active probability p_1.

currence probability of active users p_1. It is seen that the performance degrades as p_1 becomes larger. This can be explained by the fact that a larger p_1 leads to more active users in FTN-SCMA systems and both inter-user and inter-symbol interferences become severer. Also, the performance of MPA-Known algorithm with different p_1 is illustrated as performance bound. It can be observed when p_1 is small, the proposed joint estimation algorithm is capable of attaining the bound. When p_1 increases, although a small performance gap emerges, the proposed algorithm is still efficient in channel estimation.

6.6 Conclusions

In this chapter, an uplink SCMA system that utilises FTN signaling to further increase the spectral efficiency is considered. Using AR model, the correlated noise

samples are approximated by an AR process. Then based on the factorisation of the joint posterior distribution, a factor graph based hybrid message passing receiver is proposed for estimating channel coefficients and FTN data symbols. Moreover, considering a grant-free transmission scheme, the factor graph model is extended and two novel user activity detection methods are proposed. Consequently, the proposed receiver can deal with joint channel estimation, decoding and active user identification problem in FTN-SCMA systems. Simulation results show that the combined FTN-SCMA systme with the proposed receiver increases the data rate by more than 80% than the orthogonal communications systems.

Chapter 7

Conclusions

7.1 Summary of Contributions

In this thesis, the receiver design for high spectral efficiency communication system is studied. The iterative receivers for FTN signaling and SCMA system over complex channels are designed respectively. After combination of the two technologies, the spectral efficiency is further improved. This thesis has achieved innovative research results as summarised below:

1. An iterative receiver based on hybrid BP-EP-VMP algorithm is proposed for FTN signaling over unknown frequency selective channels, which solves the joint FTN symbol detection and channel estimation problem. By separating the ISI caused by FTN signaling and fading channels, the known structure of ISI induced by FTN signaling is fully exploited. A P-order AR process is used to model the colored noise so that the whitening filtering is no longer required. EP algorithm is utilised to approximate the output discrete distribution from channel decoder to a Gaussian distribution and VMP is used to tackle the message updatings at the product node. Then all messages on factor graph can be represented in Gaussian forms.

The above results have been organised and presented in [J-11].

2. Gaussian message passing based and variational inference based FDE algorithms are proposed to solve the FTN symbol detection problem over doubly selective channels. Firstly, a mean field approximation based variational inference receiver is proposed. To solve the variance underestimation

problem caused by MF approximation, Bethe approximation based receiver is developed. Then, based on the state space model of the received signal, the Forney-style factor graph is given and GMP based receiver is proposed. Finally considering the case that the channel state information is uncertain, a robust receiver is proposed by extending the factor graph.

The above results have been organised and presented in [J-2], [J-10] and [C-3].

3. An energy minimisation based receiver is designed for uplink SCMA system and its convergence behavior is analysed. The joint posterior distribution of SCMA symbols is factorised as the product of several clique potentials. Then, based on variational free energy framework, a symbol detection algorithm is proposed by minimising the corresponding free energy. Next, the convergence of the proposed algorithm is analysed. It is shown that the variance is guaranteed to convergence while the conditions that make the mean converge are derived.

The above results have been organised and presented in [J-5].

4. A convergence guaranteed message passing algorithm is proposed for downlink MIMO-SCMA system and two cooperative detection schemes are proposed for the considered multiuser system. Considering the MIMO-SCMA system, multi-user interference and ISI make the receiver very complex. For this reason, a stretched factor graph structure is designed to reduce the receiver complexity. Then, considering that BP message passing algorithm may not converge on a loopy factor graph, a convergence guaranteed message passing algorithm is designed by convexifying the Bethe free energy. To solve the cooperative detection problem in a network with noisy inter-user links, two schemes, namely, belief consensus based and ADMM based methods, are proposed.

The above results have been organised and presented in [J-3] and [C-2]

5. A joint channel estimation and data detection method is developed

for a combined FTN-SCMA system. Extended receiver structure that solves active user detection jointly is proposed for grant-free systems. Firstly, an iterative receiver based on factor graph and BP is proposed for FTN-SCMA system combining non-orthogonal waveform and non-orthogonal multiple access. Then, considering that in the grant-free system that the base station needs to detect the user's activity according to the received signal, a factor graph approach based on weighted probability is proposed. Furthermore, a unified message passing receiver is proposed that simultaneously solves the problem of active user detection, channel estimation and decoding.

The above results have been organised and presented in [J-8].

7.2 Future Work

With the rapid development of communication technologies, the following three research directions may be the focus of the future work:

1. Symbol detection in nonlinear channels. In communication systems, due to the use of high power amplifier, the nonlinearity is unavoidable. Although the digital pre-distortion technique can tackle the nonlinear distortion at the transmitter side, it is difficult to apply to practical mobile communication systems. For this reason, it is of significant importance to study the detection algorithm for a nonlinear high spectral efficiency system, for example, the FTN signaling transmitted through a nonlinear channel. Since the FTN signaling introduces ISI, the combination with channel's nonlinearity will result in a large memory nonlinear channel. Existing nonlinear channel model focuses on Volterra model. However, Volterra model has multiple orders, which makes the detection of FTN symbols difficult. The particle filtering method can deal with the symbol detection problem in nonlinear channels. However, the complexity is prohibitively high. Therefore, it is necessary to design low complexity receiver. One possible method is to use channel shortening, which

uses a post filter to shorten the channel memory depth while achieving near optimal BER performance. Nonetheless, for mobile devices, it is difficult to perform post filtering. Another way is to approximate the nonlinear channel by a linear one and then greatly reduce the complexity with negligible performance loss. How to linearise nonlinear channels is worthy of further investigation.

2. Asynchronous networks. This thesis assumes the transmitter and the receiver are fully synchronised. However, the synchronisation problem is always a key issue in wireless communications. For high spectral efficiency communication systems, due to intentional induced interferences, the carrier frequency offset and phase have greater impact. To this end, how to deal with carrier frequency offset and phase noise needs to be studied. The use of pilot sequences to track the offset and phase can be still considered, but this kind of methods degrades the spectral efficiency significantly. Therefore, it is natural to explore a joint data detection and synchronisation algorithm under one framework.

3. Active user detection in a dynamic environment. Chapter 5 in this thesis introduced active user detection method in grant-free system based on factor graph and message passing algorithm. In practice, the states of mobile users are usually time varying. An inactive user may become active in the next time slot. How to decide which users are active at the base station to achieve grant-free transmission in a dynamic environment is of significant importance. In general, we can formulate the state transition model based on empirical data to describe the activities of users at different time slots. Nevertheless, this will lead to a large number of unknown variables and result in a factor graph with many short cycles, which makes the receiver design challenging. How to find a more effective model to represent the dynamic states of users and design low complexity receivers for this system needs further study.

Appendix A

Minimizing Constrained Mean Field Free Energy

Substituting the mean field approximation $q(\mathbf{x}_f) = \prod_i q(x_{f,i})$ into (3.16) and taking into account the normalization constraint $\int q(x_{f,i})\mathrm{d}x_{f,i} = 1$, the MF VFE can be written as

$$\mathcal{F}_{MF} = -\int q(\mathbf{x}_f) \ln p(\mathbf{x}_f|\mathbf{r}_f)\mathrm{d}\mathbf{x}_f + \sum_i \int q(x_{f,i}) \ln q(x_{f,i})\mathrm{d}x_{f,i}. \tag{A.1}$$

Letting $\mu_i(x_{f,i})$ be a Lagrangian multiplier associated to the normalization constraint, the Lagrangian is given as

$$\mathcal{L}_{MF} = \mathcal{F}_{MF} + \sum_i \mu_i(x_{f,i}) \left(1 - \int q(x_{f,i})\mathrm{d}x_{f,i}\right). \tag{A.2}$$

Setting the partial derivative of Lagrangian $\nabla_{\mu_i}\mathcal{L}_{MF}$ to zero gives the normalization constraint. Computing the partial derivative of Lagrangian with respect to $q(x_{f,i})$ and setting it to zero, the approximate marginal is obtained as shown in (3.17).

Appendix B

Minimizing Constrained Bethe Free Energy

Substituting Bethe approximation (3.20) into (3.16) and taking into account the normalization constraint, the corresponding Bethe VFE can be written as

$$\begin{aligned}
\mathcal{F}_B = &- \int q(\mathbf{x}_f) \ln p(\mathbf{x}_f|\mathbf{r}_f) \mathrm{d}\mathbf{x}_f + \sum_i \int q(x_{f,i}) \ln q(x_{f,i}) \mathrm{d}x_{f,i} \\
&+ \sum_{i,j,j\neq i} \int q(x_{f,i}, x_{f,j}) \ln q(x_{f,i}, x_{f,j}) \mathrm{d}x_{f,i} \mathrm{d}x_{f,j} \\
&- \sum_{i,j,j\neq i} \int q(x_{f,i}, x_{f,j}) \ln q(x_{f,i}) q(x_{f,j}) \mathrm{d}x_{f,i} \mathrm{d}x_{f,j} \\
= &\sum_{i,j,j\neq i} \int q(x_{f,i}, x_{f,j}) \ln \frac{q(x_{f,i}, x_{f,j})}{\varphi_{ij}(x_{f,i}, x_{f,j})} \mathrm{d}x_{f,i} \mathrm{d}x_{f,j} \\
&+ (N-1) \sum_i q(x_{f,i}) \left(\ln q(x_{f,i}) - \ln \varphi_{ii}(x_{f,i}) \right). \qquad \text{(B.1)}
\end{aligned}$$

The minimization of Bethe VFE is subject to normalization constraint $\int q(x_{f,i}) \mathrm{d}x_{f,i} = 1$ and marginalization constraint $\int q(x_{f,i}, x_{f,j}) \mathrm{d}x_{f,j} = q(x_{f,i})$. Letting the associated Lagrangian multipliers being $\mu_i(x_{f,i})$ and $\mu_{ij}(x_{f,i})$, the Lagrangian of Bethe approximation is given as

$$\begin{aligned}
\mathcal{L}_B = \mathcal{F}_B &+ \sum_{i,j,j\neq i} \mu_{ij}(x_{f,i}) \left(q(x_{f,i}) - \int q(x_{f,i}, x_{f,j}) \mathrm{d}x_{f,j} \right) \\
&+ \sum_i \mu_i(x_{f,i}) \left(1 - \int q(x_{f,i}) \mathrm{d}x_{f,i} \right). \qquad \text{(B.2)}
\end{aligned}$$

Setting $\nabla_{\mu_i(x_{f,i})} \mathcal{L}_B = 0$ and $\nabla_{\mu_{ij}(x_{f,i})} \mathcal{L}_B = 0$ results in the normalization and marginalization constraints. Setting the partial derivatives $\nabla_{q(x_{f,i})} \mathcal{L}_B$ and $\nabla_{q(x_{f,i}, x_{f,j})} \mathcal{L}_B$

to zero yields

$$\ln q(x_{f,i}) = \mu_i(x_{f,i}) + \ln \varphi_{ii}(x_{f,i}) + \sum_{j,j\neq i} \mu_{ij}(x_{f,i}), \tag{B.3}$$

$$\ln q(x_{f,i}, x_{f,j}) = \varphi_{ij}(x_{f,i}, x_{f,j}) + \ln q(x_{f,i})q(x_{f,j}) - \mu_{ij}(x_{f,i}) - \mu_{ji}(x_{f,j}). \tag{B.4}$$

Substituting (B.3) into (B.4) and taking exponential function on both sides, yields

$$q(x_{f,i}, x_{f,j}) = \mathrm{C}_1 \varphi_{ii}(x_{f,i})\varphi_{jj}(x_{f,j})\varphi_{ij}(x_{f,i}, x_{f,j}) \prod_{k,k\neq i,k\neq j} \exp(\mu_{ik}(x_{f,i})) \exp(\mu_{jk}(x_{f,j})) \tag{B.5}$$

where C_1 is a constant. Integrating $x_{f,j}$ from above equation and comparing it with (B.3) yield

$$\exp(\mu_{ij}(x_{f,i})) \propto \int \varphi_{ij}(x_{f,i}, x_{f,j}) \prod_{\substack{k,k\neq i \\ k\neq j}} \exp(\mu_{jk}(x_{f,j}))\mathrm{d}x_{f,j}.$$

Letting $\tilde{q}_{\setminus i}(x_{f,j}) = \prod_{k,k\neq i,k\neq j} \exp(\mu_{jk}(x_{f,j}))$ and substituting it into (B.3), the approximate marginal is obtained as shown in (3.21).

Appendix C

Proof of Proposition 1

According to (4.18) and (4.24), the update equation of $\beta_{j,i}^{k}(l)$ is rewritten as

$$\beta_{K,i}^{j}(l) = -a\left(\rho_i^J + \sum_{(i,m),m\neq k} \beta_{i,m}^{j}(l-1)\right)^{-1}, \tag{C.1}$$

with $a = h_{k,j}^4/N_0^2$. Note that if $\beta_{i,k}^{j} \leq 0$, the following inequality can be derived

$$\begin{aligned}\beta_{k,i}^{j}(l+1) - \beta_{k,i}^{j}(l) =& \frac{a\sum_{(i,m),m\neq k}\left[\beta_{i,m}^{j}(l) - \beta_{i,m}^{j}(l-1)\right]}{\left(\rho_i^j + \sum_{(i,m),m\neq k}\beta_{i,m}^{j}(l-1)\right)\left(\rho_i^j + \sum_{(i,m),m\neq k}\beta_{i,m}^{j}(l)\right)} \\ &\geq \frac{a}{(\rho_i^j)^2}\sum_{(i,m),m\neq k}\left[\beta_{i,m}^{j}(l) - \beta_{i,m}^{j}(l-1)\right]. \end{aligned}\tag{C.2}$$

By stacking all β values with respect to the resource index j to form $\boldsymbol{\beta}^j$, the above inequality can be expressed in a vectorial form as

$$\begin{aligned}\boldsymbol{\beta}^j(l+1) - \boldsymbol{\beta}^j(l) &\geq \frac{a}{(\rho_i^j)^2}\mathbf{A}\left[\boldsymbol{\beta}^j(l) - \boldsymbol{\beta}^j(1-1)\right] \\ &\geq \frac{a^l}{(\rho_i^j)^{2l}}\mathbf{A}^l\left[\boldsymbol{\beta}^j(1) - \boldsymbol{\beta}^j(0)\right], \end{aligned}\tag{C.3}$$

where $\mathbf{A}$ is an adjacent matrix with $\mathbf{A}_{ik} = 1$ if and only if user i and k are interfering with each other. Since the symbols are initialized with $v_{x_{k,j}}(0) = \infty$, which indicates that $v_{x_{k,j}}(1) \leq v_{x_{k,j}}(0)$, therefore $\beta_{k,i}^{j}(1) \geq \beta_{k,i}^{j}(0)$ holds and $\boldsymbol{\beta}^j(l+1) - \boldsymbol{\beta}^j(l) \geq \mathbf{0}$ is obtained, which shows that the parameter $\beta_{k,i}^{j}(l+1)$ is monotonically increasing. According to (4.20), $v_{x_{k,j}}(l+1) < v_{x_{k,j}}(l)$, which proves that $v_{x_{k,j}}$ is guaranteed to converge.

Considering the belief damping, to prove $\tilde{v}_{x_{k,j}}(l+1) < \tilde{v}_{x_{k,j}}(l)$ is equivalent to prove

$$\frac{\alpha}{v_{x_{k,j}}(l)} + \frac{1-\alpha}{v_{x_{k,j}}(l+1)} \geq \frac{\alpha}{v_{x_{k,j}}(l-1)} + \frac{1-\alpha}{v_{x_{k,j}}(l)}. \tag{C.4}$$

Obviously, with the conclusion drawn above, $\frac{\alpha}{v_{x_{k,j}}(l)} \geq \frac{\alpha}{v_{x_{k,j}}(l-1)}$ and $\frac{1-\alpha}{v_{x_{k,j}}(l+1)} \geq \frac{1-\alpha}{v_{x_{k,j}}(l)}$ always hold. Therefore the variance of damped belief is also guaranteed to converge.

Appendix D

Proof of Proposition 2

• *Necessary Condition*: Based on (4.32), $\boldsymbol{\gamma}^j(l+1) - \boldsymbol{\gamma}^j(l)$ can be expressed as

$$\boldsymbol{\gamma}^j(l+1) - \boldsymbol{\gamma}^j(l) = b^l \mathbf{A}^l \left[\boldsymbol{\gamma}^j(1) - \boldsymbol{\gamma}^j(0)\right]. \tag{D.1}$$

Deriving the limits of both sides of (D.1) yields

$$\lim_{l\to\infty} \left[\boldsymbol{\gamma}^j(l+1) - \boldsymbol{\gamma}^j(l)\right] = \lim_{l\to\infty} (b\mathbf{A})^l \left[\boldsymbol{\gamma}^j(1) - \boldsymbol{\gamma}^j(0)\right]. \tag{D.2}$$

If $\tilde{m}_{x_{k,j}}$ converges, we have

$$\begin{aligned} &\lim_{l\to\infty} \left[\alpha\left(\boldsymbol{\gamma}^j(l+1) - \boldsymbol{\gamma}^j(l)\right) + (1-\alpha)\left(\boldsymbol{\gamma}^j(l) - \boldsymbol{\gamma}^j(l-1)\right)\right] \\ &= \lim_{l\to\infty} \left[\alpha(b\mathbf{A})^l + (1-\alpha)(b\mathbf{A})^{l-1}\right] \cdot \left[\boldsymbol{\gamma}^j(1) - \boldsymbol{\gamma}^j(0)\right] = 0 \end{aligned} \tag{D.3}$$

which in turn requires $\lim_{l\to\infty}(b\mathbf{A})^l = 0$. Assuming that λ and $\boldsymbol{\nu}$ are the eigenvalue and eigenvector of $\mathbf{A}$, then

$$\begin{aligned} \boldsymbol{\nu} \lim_{l\to\infty} (b\mathbf{A})^l &= \lim_{l\to\infty} \left[(b\mathbf{A})^l \boldsymbol{\nu}\right] = \lim_{l\to\infty} \left[(b\lambda)^l \boldsymbol{\nu}\right] \\ &= \boldsymbol{\nu} \lim_{l\to\infty} (b\lambda)^l. \end{aligned} \tag{D.4}$$

Since $\boldsymbol{\nu} \neq \mathbf{0}$, $\lim_{l\to\infty}(b\lambda)^l = 0$ must be satisfied. Therefore $|\lambda| < \frac{1}{b}$ holds for any eigenvalue of the matrix $\mathbf{A}$. Consequently, $\rho(\mathbf{A}) < \frac{1}{b}$.

• *Sufficient Condition*: According to the matrix theorem of [134], $\|\boldsymbol{\gamma}^j(l+1) - \boldsymbol{\gamma}^j(l)\|$ can be written as

$$\begin{aligned} \|\boldsymbol{\gamma}^j(l+1) - \boldsymbol{\gamma}^j(l)\| &= \|b\mathbf{A}\left[\boldsymbol{\gamma}^j(l) - \boldsymbol{\gamma}^j(l-1)\right]\| \\ &\leq b \cdot \rho(\mathbf{A}) \|\boldsymbol{\gamma}^j(l) - \boldsymbol{\gamma}^j(l-1)\|. \end{aligned} \tag{D.5}$$

If $\rho(\mathbf{A}) < \frac{1}{b}$, then $\|\boldsymbol{\gamma}^j(l+1) - \boldsymbol{\gamma}^j(l)\| < \|\boldsymbol{\gamma}^j(l) - \boldsymbol{\gamma}^j(l-1)\|$ holds and therefore $\tilde{m}_{x_{k,j}}$ is convergent.

Appendix E

Derivation of Messages (5.41) and (5.42)

Solving the optimization problem (5.40) yields the corresponding beliefs as

$$b_a(\mathbf{x}_a) \propto f_a(\mathbf{x}_a)^{\frac{1}{c_a}} \prod_{i \in \mathcal{N}(a)} \exp(\frac{\beta_{ai}(x_i)}{c_a}) \tag{E.1}$$

$$b_i(x_i) \propto \prod_{f_a \in \mathcal{N}(i)} \exp(-\frac{\beta_{ai}(x_i)}{c_i}). \tag{E.2}$$

For clarity, the following definitions are made: $\tau_i = (1 - c_i)/|\mathcal{N}(i)|$, $\mu_{x_i \to f_a}(x_i) = \exp(\frac{\beta_{ai}(x_i)}{c_a})$ and $\mu_{f_a \to x_i}(x_i) = b_i^{\tau_i}(x_i) \exp(-\beta_{ai}(x_i))$. Then the following equations are obtained

$$\exp(\beta_{ai}(x_i)) = \mu_{x_i \to f_a}^{c_a}(x_i) \tag{E.3}$$

$$\exp(-\beta_{ai}(x_i)) = \mu_{f_a \to x_i}(x_i) \cdot b_i^{-\tau_i}(x_i). \tag{E.4}$$

Substituting (E.3) and (E.4) into (E.1) and (E.2) yields,

$$b_a(\mathbf{x}_a) \propto f_a(\mathbf{x}_a)^{\frac{1}{c_a}} \prod_{i \in \mathcal{N}(a)} \mu_{x_i \to f_a}(x_i) \tag{E.5}$$

$$b_i^{c_i + \tau_i |\mathcal{N}(i)|}(x_i) \propto \prod_{f_a \in \mathcal{N}(i)} \mu_{f_a \to x_i}(x_i). \tag{E.6}$$

Define two auxiliary messages as

$$\tilde{\mu}_{f_a \to x_i}(x_i) \propto \int f_a^{\frac{1}{c_a}}(\mathbf{x}_a) \prod_{i' \in \mathcal{N}(a) \setminus i} \mu_{x_{i'} \to f_a}(x_{i'}) \mathrm{d}x_{i'} \tag{E.7}$$

$$\tilde{\mu}_{x_i \to f_a}(x_i) \propto \prod_{f_{a'} \in \mathcal{N}(i) \setminus f_a} \mu_{f_{a'} \to i}(x_i). \tag{E.8}$$

From (E.5)-(E.8), we have

$$\begin{aligned} b_i(x_i) &= \tilde{\mu}_{x_i \to f_a}(x_i)\mu_{f_a \to x_i}(x_i) \\ &= \int b_a(\mathbf{x}_a)\mathrm{d}\mathbf{x}_a \backslash x_i \\ &= \tilde{\mu}_{f_a \to x_i}(x_i)\mu_{x_i \to f_a}(x_i). \end{aligned} \tag{E.9}$$

Comparing (E.9) to $\mu^{c_a}_{x_i \to f_a}(x_i)\mu_{f_a \to x_i}(x_i) = b_i^{\tau_i}(x_i)$ gives

$$\mu^{c_a}_{x_i \to f_a}(x_i)\mu_{f_a \to x_i}(x_i) = \tilde{\mu}^{\tau_i}_{x_i \to f_a}(x_i)\mu^{\tau_i}_{f_a \to x_i}(x_i) \tag{E.10}$$

and then

$$\mu_{x_i \to f_a}(x_i) = \tilde{\mu}^{\frac{\tau_i}{c_a}}_{x_i \to f_a}(x_i)\mu^{\frac{\tau_i - 1}{c_a}}_{f_a \to x_i}(x_i). \tag{E.11}$$

Based on (E.9) and (E.11), we have

$$\mu^{\frac{c_a - \tau_i + 1}{c_a}}_{f_a \to x_i}(x_i) = \tilde{\mu}_{f_a \to x_i}(x_i)\tilde{\mu}^{\frac{\tau_i - c_a}{c_a}}_{x_i \to f_a}(x_i) \tag{E.12}$$

$$\mu_{f_a \to x_i}(x_i) = \tilde{\mu}^{\frac{\tau_i - c_a}{c_a - \tau_i + 1}}_{x_i \to f_a}(x_i)\tilde{\mu}^{\frac{c_a}{c_a - \tau_i + 1}}_{f_a \to x_i}(x_i). \tag{E.13}$$

Finally, substituting (E.13) into (E.11) yields

$$\mu_{x_i \to f_a}(x_i) = \tilde{\mu}^{\frac{1}{c_a - \tau_i + 1}}_{x_i \to f_a}(x_i)\tilde{\mu}^{\frac{\tau_i - 1}{c_a - \tau_i + 1}}_{f_a \to x_i}(x_i). \tag{E.14}$$

With the definition of τ_i, it is easy to see messages (E.13) and (E.14) are the same as (5.41) and (5.42).

Appendix F

Derivations of (6.48) and (6.49)

For the product vertex $\times_{kj}^n$, the message $\mu_{\bar{s}_{kj}^n \to f_j^n}$ can be regarded as the distribution of $\bar{s}_{kj}^n = \xi_k \tilde{s}_{kj}^n$ with random variables ξ_k and $\tilde{s}_{kj}^n$ following distributions $\mu_{\xi_k \to \times_{kj}^n}$ and $\mu_{\tilde{s}_{kj}^n \to \times_{kj}^n}$. Since $\mu_{\xi_k \to \times_{kj}^n}$ and $\mu_{\tilde{s}_{kj}^n \to \times_{kj}^n}$ are both Gaussian distribution, the density of $\bar{s}_{kj}^n$ is calculated as

$$\begin{aligned} f(\bar{s}_{kj}^n) &= \int f(\tilde{s}_{kj}^n) f(\xi_k) \delta(\bar{s}_{kj}^n - \tilde{s}_{kj}^n \xi_k) \mathrm{d}\tilde{s}_{kj}^n \mathrm{d}\xi_k \\ &= \int \frac{1}{|\xi_k|} f\left(\frac{\bar{s}_{kj}^n}{\xi_k}\right) f(\xi_k) \mathrm{d}\xi_k \\ \propto \int \frac{1}{|\xi_k|} \exp&\left(- \frac{(\frac{\bar{s}_{kj}^n}{\xi_k} - m_{\tilde{s}_{kj}^n \to \times_{kj}^n})^2}{v_{\tilde{s}_{kj}^n \to \times_{kj}^n}} - \frac{(\xi_k - m_{\xi_k \to \times_{kj}^n})^2}{v_{\xi_k \to \times_{kj}^n}} \right) \mathrm{d}\xi_k. \end{aligned} \tag{F.1}$$

However, the above integral does not have a analytical expression. The goal is to derive a Gaussian message, that is to determine the mean and variance of $\mu_{\bar{s}_{kj}^n \to f_j^n}$ based on incoming messages.

It is well known for two independent random variables x and y, based on Mellin Transform [135], the nth-order moment of xy satisfies

$$\mathbb{E}[(xy)^n] = \mathbb{E}(x^n)\mathbb{E}(y^n). \tag{F.2}$$

Thus the first two order moments of $\mu_{\bar{s}_{kj}^n \to f_j^n}$ are given as

$$\mathbb{E}[\bar{s}_{kj}^n] = \mathbb{E}[\tilde{s}_{kj}^n]\mathbb{E}[\xi_k] = m_{\xi_k \to \times_{kj}^n} m_{\phi_{kj}^n \to \tilde{s}_{kj}^n}, \tag{F.3}$$

$$\begin{aligned} \mathbb{E}[(\bar{s}_{kj}^n)^2] &= \mathbb{E}[(\tilde{s}_{kj}^n)^2]\mathbb{E}[\xi_k^2] \\ &= (m_{\xi_k \to \times_{kj}^n}^2 + v_{\xi_k \to \times_{kj}^n})(m_{\phi_{kj}^n \to \bar{s}_{kj}^n}^2 + v_{\phi_{kj}^n \to \bar{s}_{kj}^n}), \end{aligned} \tag{F.4}$$

and the variance $v_{\bar{s}_{kj}^n \to f_j^n} = \mathbb{E}[(\bar{s}_{kj}^n)^2] - \mathbb{E}^2[\bar{s}_{kj}^n]$, which are given as (6.48) and (6.49).

Bibliography

[1] W. R. Young, "Advanced mobile phone service: introduction, background, and objectives," *Bell System Technical Journal*, vol. 58, no. 1, pp. 1–14, 1979.

[2] M. Mouly, M.-B. Pautet, and T. Foreword By-Haug, *The GSM system for mobile communications*. Telecom publishing, 1992.

[3] D. D. Falconer, F. Adachi, and B. Gudmundson, "Time division multiple access methods for wireless personal communications," *IEEE Commun. Mag.*, vol. 33, no. 1, pp. 50–57, 1995.

[4] A. J. Viterbi, *CDMA: principles of spread spectrum communication*. Addison-Wesley Reading, MA, 1995, vol. 122.

[5] D. Martín-Sacristán, J. F. Monserrat, J. Cabrejas-Peñuelas, D. Calabuig, S. Garrigas, and N. Cardona, "On the way towards fourth-generation mobile: 3GPP LTE and LTE-advanced," *EURASIP J. Wireless Commun., Network.*, vol. 2009, no. 1, p. 354089, 2009.

[6] H. Holma and A. Toskala, *LTE for UMTS: OFDMA and SC-FDMA based radio access*. John Wiley & Sons, 2009.

[7] H. Weingarten, Y. Steinberg, and S. S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, 2006.

[8] R. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, 1962.

[9] A. Varghese and D. Tandur, "Wireless requirements and challenges in industry 4.0," in *Proc. Int. Conf. Contemporary Computing and Informatics (IC3I)*, 2014, pp. 634–638.

[10] J. Thompson, X. Ge, H.-C. Wu, R. Irmer, H. Jiang, G. Fettweis, and S. Alamouti, "5g wireless communication systems: Prospects and challenges," *IEEE Commun. Mag*, vol. 52, no. 2, pp. 62–64, 2014.

[11] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (IoT): A vision, architectural elements, and future directions," *Future Gen. Computer Sys.*, vol. 29, no. 7, pp. 1645–1660, 2013.

[12] T. Xu, J. B. Wendt, and M. Potkonjak, "Security of IoT systems: Design challenges and opportunities," in *Proc. the IEEE/ACM Int. Conf. Computer-Aided Design*, 2014, pp. 417–423.

[13] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, 2016.

[14] R. Dou and G. Nan, "Optimizing sensor network coverage and regional connectivity in industrial IoT systems," *IEEE Sys. J.*, vol. 11, no. 3, pp. 1351–1360, 2017.

[15] N. Alliance, "5G white paper," *Next generation mobile networks, white paper*, pp. 1–125, 2015.

[16] C. G, F. T, M. A, and P. A, "Faster-than-Nyquist and beyond: how to improve spectral efficiency by accepting interference." *Optics Express*, vol. 19, no. 27, pp. 26 600–26 609, 2011.

[17] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen,

S. Ten Brink, I. Gaspar, N. Michailow, A. Festag *et al.*, "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 97–105, 2014.

[18] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, 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] Z. Gao, L. Dai, D. Mi, and Z. Wang, "Mmwave massive-mimo-based wireless backhaul for the 5g ultra-dense network," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 13–21, 2015.

[21] W. Hong, K. H. Baek, Y. Lee, and Y. Kim, "Study and prototyping of practically large-scale mmwave antenna systems for 5g cellular devices," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 63–69, 2014.

[22] A. Jovicic, J. Li, and T. Richardson, "Visible light communication: opportunities, challenges and the path to market," *IEEE Commun. Mag.*, vol. 51, no. 12, pp. 26–32, 2013.

[23] M. Ji, G. Caire, and A. F. Molisch, "Fundamental limits of caching in wireless d2d networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 849–869, 2016.

[24] G. Wu, S. Talwar, K. Johnsson, and N. Himayat, "M2m: From mobile to embedded internet," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 36–43, 2011.

[25] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, 2010.

[26] B. Farhang-Boroujeny, "Ofdm versus filter bank multicarrier," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 92–112, 2011.

[27] F. Schaich and T. Wild, "Waveform contenders for 5g ofdm vs. fbmc vs. ufmc," in *Proc. Int. Symp. Commun., Control, Signal Process.*

[28] J. E. Mazo, "Faster-than-nyquist signaling," *Bell Labs Tech. J.*, vol. 54, no. 8, pp. 1451–1462, 1975.

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

[30] N. Benvenuto, D. Rui, D. Falconer, and S. Tomasin, "Single carrier modulation with nonlinear frequency domain equalization: An idea whose time has comeagain," *Proc. IEEE*, vol. 98, no. 1, pp. 69–96, 2009.

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

[32] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, 2015.

[33] N. Zhang, J. Wang, G. Kang, and Y. Liu, "Uplink nonorthogonal multiple access in 5g systems," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 458–461, 2016.

[34] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. S. Kwak, "Power-domain non-orthogonal multiple access (noma) in 5g systems: Potentials and challenges," *IEEE Commun. Surveys & Tut.*, vol. PP, no. 99, pp. 1–1, 2016.

[35] L. Ping, L. Liu, K. Wu, and W. K. Leung, "Interleave division multiple-access," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 938–947, 2006.

[36] C. Novak, F. Hlawatsch, and G. Matz, "Mimo-idma: Uplink multiuser mimo communications using interleave-division multiple access and low-complexity iterative receivers," in *Proc. IEEE Int. Conf. Acous., Speech, Signal Process.*, 2007, pp. III–225 – III–228.

[37] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for Internet of things," in *Proc. IEEE Veh. Technol. Conf.*, 2016, pp. 1–5.

[38] J. Zeng, B. Li, X. Su, L. Rong, and R. Xing, "Pattern division multiple access (pdma) for cellular future radio access," in *Proc. Int. Conf. Wireless Commun., Signal Process.*, 2015, pp. 1–5.

[39] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE 24th Int. Symp. Person. Indoor, Mobile Radio Commun.*, 2013, pp. 332–336.

[40] S. Zhang, X. Xu, L. Lu, Y. Wu, G. He, and Y. Chen, "Sparse code multiple access: An energy efficient uplink approach for 5g wireless systems," in *Proc. Global Commun. Conf.*, 2014, pp. 4782–4787.

[41] R. Lucky, "Decision feedback and faster-than-nyquist transmission," in *Proc. IEEE Int. Sympo. Inf. Theory*, 1970, pp. 15–19.

[42] T. S. Rappaport *et al.*, *Wireless communications: principles and practice.* prentice hall PTR New Jersey, 1996, vol. 2.

[43] F. Ling, "Matched filter-bound for time-discrete multipath rayleigh fading channels," *IEEE Trans. Commun.*, vol. 43, no. 2, pp. 710–713, 1995.

[44] C. K. Wang and L. S. Lee, "Practically realizable digital transmission significantly below the nyquist bandwidth," in *Proc. Global Telecommun. Conf.*, 1991, pp. 1187–1191 vol.2.

[45] R. D. Cideciyan, J. D. Coker, E. Eleftheriou, and R. L. Galbraith, "Noise predictive maximum likelihood detection combined with parity-based post-processing," *IEEE Trans. Mag.*, vol. 37, no. 2, pp. 714–720, 2001.

[46] A. J. Viterbi, *Viterbi Algorithm.* John Wiley & Sons, Inc., 2003.

[47] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Tran. Inf. Theory*, vol. 20, no. 2, pp. 284–287, 1974.

[48] N. Al-Dhahir, "Fir channel-shortening equalizers for mimo isi channels," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 213–218, 2001.

[49] 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, 1998.

[50] M. Eyuboglu and S. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. 36, no. 1, pp. 13–20, 1988.

[51] J. B. Anderson and A. Prlja, "Turbo equalization and an m-bcjr algorithm for strongly narrowband intersymbol interference," in *Proc. Int. Symp. Inf. Theory*, 2010, pp. 261–266.

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

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

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

[55] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[56] M. Moltafet, N. M. Yamchi, M. R. Javan, and P. Azmi, "Comparison study between pd-noma and scma," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1830–1834, 2017.

[57] Y. Wu, S. Zhang, and Y. Chen, "Iterative multiuser receiver in sparse code multiple access systems," in *Proc. IEEE Int. Conf. Commun.*, 2015, pp. 2918–2923.

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

[59] Y. Du, B. Dong, Z. Chen, J. Fang, and L. Yang, "Shuffled multiuser detection schemes for uplink sparse code multiple access systems," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1231–1234, June 2016.

[60] H. Mu, Z. Ma, M. Alhaji, P. Fan, and D. Chen, "A fixed low complexity message pass algorithm detector for up-link SCMA system," *IEEE Wireless Commun. Lett.*, vol. 4, no. 6, pp. 585–588, 2015.

[61] J. Chen, Z. Zhang, S. He, J. Hu, and G. E. Sobelman, "Sparse code multiple access decoding based on a Monte Carlo Markov chain method," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 639–643, 2016.

[62] F. Wei and W. Chen, "Low complexity iterative receiver design for sparse

code multiple access," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 621–634, Feb 2017.

[63] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: an empirical study," in *Proc. 15th Conf. Uncertain. Artif. Intell.*, 1999, pp. 467–475.

[64] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 67–75, 2002.

[65] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 223–229, 2002.

[66] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1207–1215, 2000.

[67] Y.-S. Choi, P. J. Voltz, and F. A. Cassara, "On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 49, no. 8, pp. 1375–1387, Aug 2001.

[68] C. Cozzo and B. L. Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1266–1270, Aug 2003.

[69] F. Li, Z. Xu, and S. Zhu, "Variational-inference-based data detection for OFDM systems with imperfect channel estimation," *IEEE Trans. Veh. Technol.*, vol. 62, no. 3, pp. 1394–1399, Mar 2013.

[70] H. Niu, M. Shen, J. A. Ritcey, and H. Liu, "A factor graph approach to iterative channel estimation and LDPC decoding over fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1345–1350, 2005.

[71] G. E. Kirkelund, C. N. Manchón, L. P. Christensen, E. Riegler, and B. H. Fleury, "Variational message-passing for joint channel estimation and decoding in MIMO-OFDM," in *Proc. IEEE Global Telecommun. Conf.*, 2010, pp. 1–6.

[72] S. Wu, L. Kuang, Z. Ni, J. Lu, D. D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE J. Select. Topics Signal Process.*, vol. 8, no. 5, pp. 902–915, 2014.

[73] 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.

[74] D. Dasalukunte, V. wall, F. Rusek, and J. B. Anderson, *Faster than Nyquist signaling: Algorithms to silicon.* Springer Netherlands, 2014.

[75] J. D. Gibson, B. Koo, and S. D. Gray, "Filtering of colored noise for speech enhancement and coding," *IEEE Trans. Signal Process.*, vol. 39, no. 8, pp. 1732–1742, 1991.

[76] 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, 2007.

[77] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2840–2849, 2014.

[78] J. M. Winn and C. M. Bishop, "Variational message passing," *J. Machine Learn. Res.*, pp. 661–694, 2005.

[79] J. Dauwels, A. Eckford, S. Korl, and H.-A. Loeliger, "Expectation maximization as message passing-part I: Principles and Gaussian messages," *arXiv preprint arXiv:0910.2832*, 2009.

[80] Q. Guo and D. D. Huang, "A concise representation for the soft-in soft-out LMMSE detector." *IEEE Commun. Lett.*, vol. 15, no. 5, pp. 566–568, 2011.

[81] 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, 2007.

[82] P. Schniter, "Low-complexity equalization of OFDM in doubly selective channels," *IEEE Tran. Signal Process.*, vol. 52, no. 4, pp. 1002–1011, 2004.

[83] X. Huang and H.-C. Wu, "Robust and efficient intercarrier interference mitigation for OFDM systems in time-varying fading channels," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2517–2528, 2007.

[84] T. Hrycak, S. Das, G. Matz, and H. G. Feichtinger, "Low complexity equalization for doubly selective channels modeled by a basis expansion," *IEEE Tran. Signal Process.*, vol. 58, no. 11, pp. 5706–5719, 2010.

[85] P. Cheng, Z. Chen, Y. Rui, Y. J. Guo, L. Gui, M. Tao, and Q. Zhang, "Channel estimation for OFDM systems over doubly selective channels: A distributed compressive sensing based approach," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4173–4185, 2013.

[86] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 255–271, April 2010.

[87] Q. Guo, L. Ping, and D. Huang, "A low-complexity iterative channel estimation and detection technique for doubly selective channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, 2009.

[88] M. Opper and D. Saad, *Advanced mean field methods: Theory and practice*. MIT press, 2001.

[89] S. Katsura and M. Takizawa, "Bethe lattice and the Bethe approximation," *Progress of Theoretical Physics*, vol. 51, no. 1, pp. 82–98, 1974.

[90] 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.

[91] G. Colavolpe and A. Barbieri, "On MAP symbol detection for ISI channels using the Ungerboeck observation model," *IEEE Commun. Lett.*, vol. 9, no. 8, pp. 720–722, 2005.

[92] F. Rusek, G. Colavolpe, and C. E. W. Sundberg, "40 years with the Ungerboeck model: A look at its potentialities [lecture notes]," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 156–161, 2015.

[93] 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, 2011.

[94] Yedidia, Jonathan S and Freeman, William T and Weiss, Yair, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2282–2312, 2005.

[95] J. Zhou, J. Qin, and Y.-C. Wu, "Variational inference-based joint interference mitigation and OFDM equalization under high mobility," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1970–1974, Nov 2015.

[96] J. Dai, K. Niu, C. Dong, and J. Lin, "Improved message passing algorithms for sparse code multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 9986–9999, Nov 2017.

[97] X. Meng, Y. Wu, Y. Chen, and M. Cheng, "Low complexity receiver for uplink SCMA system via expectation propagation," in *Proc. IEEE Wireless Commun. Networking Conf.*, 2017, pp. 1–5.

[98] G. E. Box and G. C. Tiao, *Bayesian inference in statistical analysis*. John Wiley & Sons, 2011.

[99] F. Ricci-Tersenghi, "The Bethe approximation for solving the inverse Ising problem: a comparison with other inference methods," *J. Statis. Mechan.: Theory, Experiment*, vol. 2012, no. 08, pp. 1–23, Aug 2012.

[100] 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.

[101] K. Friston, J. Mattout, N. Trujillo-Barreto, J. Ashburner, and W. Penny, "Variational free energy and the Laplace approximation," *Neuroimage*, vol. 34, no. 1, pp. 220–234, 2007.

[102] P. Som, T. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, "Low-complexity detection in large-dimension MIMO-ISI channels using graphical models," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1497–1511, Aug 2011.

[103] S. Zhang, K. Xiao, B. Xiao, Z. Chen, B. Xia, D. Chen, and S. Ma, "A capacity-based codebook design method for sparse code multiple access systems," in *Proc. 8th Int. Conf. Wireless Commun. Signal Process.*, 2016, pp. 1–5.

[104] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA codebook design," in *Proc. IEEE 80th Veh. Technol. Conf*, 2014, pp. 1–5.

[105] S. Tang, L. Hao, and Z. Ma, "Low complexity joint MPA detection for downlink MIMO-SCMA," in *Proc. IEEE Global Commun. Conf.*, Dec 2016, pp. 1–4.

[106] L. Liu, C. Yuen, Y. L. Guan, Y. Li, and C. Huang, "Gaussian message passing iterative detection for MIMO-NOMA systems with massive access," *arXiv preprint arXiv:1607.00800*, 2016.

[107] T. Wo, J. C. Fricke, and P. A. Hoeher, "A graph-based iterative gaussian detector for frequency-selective MIMO channels," in *Proc. IEEE Inf. Theory Workshop*, 2006, pp. 581–585.

[108] W. Haselmayr, B. Etzlinger, and A. Springer, "Factor-graph-based soft-input soft-output detection for frequency-selective MIMO channels," *IEEE Commun. Lett.*, vol. 16, no. 10, pp. 1624–1627, 2012.

[109] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proc. IEEE*, vol. 97, no. 2, pp. 427–450, 2009.

[110] F. Meyer, O. Hlinka, H. Wymeersch, E. Riegler, and F. Hlawatsch, "Distributed localization and tracking of mobile networks including noncooperative objects," *IEEE Trans. Signal Inf. Process. Net.*, vol. 2, no. 1, pp. 57–71, 2016.

[111] H. Zhu, A. Cano, and G. B. Giannakis, "Distributed consensus-based demodulation: algorithms and error analysis," *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 2044–2054, 2010.

[112] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, 2008.

[113] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, 2015.

[114] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[115] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals Math. Stat.*, vol. 22, no. 1, pp. 79–86, 1951.

[116] 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. 5, pp. 6849–6858, 2017.

[117] J. S. Yedidia, W. T. Freeman, Y. Weiss *et al.*, "Generalized belief propagation," in *Proc. NIPS*, vol. 13, 2000, pp. 689–695.

[118] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1568–1583, 2006.

[119] T. Heskes, "Convexity arguments for efficient minimization of the Bethe and Kikuchi free energies." *J. Artif. Intell. Res.(JAIR)*, vol. 26, pp. 153–190, 2006.

[120] L. Xiao, S. Boyd, and S. Lall, "Distributed average consensus with time-varying metropolis weights," *Automatica*, 2006.

[121] C. Song, S. Yoon, and V. Pavlovic, "Fast ADMM algorithm for distributed optimization with adaptive penalty," *arXiv preprint arXiv:1506.08928*, 2015.

[122] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Prentice hall Englewood Cliffs, NJ, 1989, vol. 23.

[123] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel low-density signature for synchronous CDMA systems over AWGN channel," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1616–1626, 2008.

[124] T. P. Minka, "Expectation propagation for approximate bayesian inference," in *Proc. Conf. Uncertainty Artif. Intell.*, 2001, pp. 362–369.

[125] G. Szabo, D. Orincsay, B. P. Gero, S. Gyori, and T. Borsos, "Traffic analysis of mobile broadband networks," in *Proc. 3rd Int. Conf. Wireless Internet*, 2007, p. 18.

[126] J. Zhang, L. Lu, Y. Sun, Y. Chen, J. Liang, J. Liu, H. Yang, S. Xing, Y. Wu, J. Ma *et al.*, "PoC of SCMA-based uplink grant-free transmission in UCNC for 5G," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1353–1362, 2017.

[127] B. Shim and B. Song, "Multiuser detection via compressive sensing," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 972–974, 2012.

[128] B. Wang, L. Dai, Y. Yuan, and Z. Wang, "Compressive sensing based multi-user detection for uplink grant-free non-orthogonal multiple access," in *Proc. IEEE 82nd Veh. Technol. Conf.*, 2015, pp. 1–5.

[129] C. Wei, H. Liu, Z. Zhang, J. Dang, and L. Wu, "Approximate message passing-based joint user activity and data detection for NOMA," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 640–643, 2017.

[130] Y. Zhang, Q. Guo, Z. Wang, J. Xi, and N. Wu, "Block sparse bayesian learning based joint user activity detection and channel estimation for grant-free NOMA systems," *IEEE Trans. Veh. Technol.*, pp. 1–1, 2018.

[131] F. Wei and W. Chen, "Message passing receiver design for uplink grant-free SCMA," in *Proc. 2017 IEEE Globecom Workshops (GC Wkshps)*, Dec 2017, pp. 1–6.

[132] W. Yuan, N. Wu, Q. Guo, Y. Li, C. Xing, and J. Kuang, "Iterative receivers for downlink MIMO-SCMA: Message passing and distributed cooperative detection," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3444–3458, 2018.

[133] B. Wang, L. Dai, Y. Zhang, T. Mir, and J. Li, "Dynamic compressive sensing-based multi-user detection for uplink grant-free NOMA," *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2320–2323, 2016.

[134] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 2012.

[135] A. M. Mathai, *A handbook of generalized special functions for statistical and physical sciences*. Oxford University Press, USA, 1993.