THE UNIVERSITY OF
SYDNEY

THE UNIVERSITY OF SYDNEY

POLAR CODING SCHEMES FOR COOPERATIVE TRANSMISSION SYSTEMS

By

Bin DUO

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
AT
SCHOOL OF ELECTRICAL AND INFORMATION ENGINEERING
FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGIES
THE UNIVERSITY OF SYDNEY

March 2016

Acknowledgements

It seems that only in the blink of an eye, my years of doctoral studies sped by. At the completion of this paper, I cannot help feeling a strong sense of fulfilment. Meanwhile, many emotional and grateful thoughts are welling up in me. My deepest gratitude goes first and foremost to Professor Xuemai Gu, who dedicated meticulous care and attention to me during my doctoral studies. I also must thank him for giving me the once-in-a-lifetime opportunity of being recommended to pursue further education at to the world-renowned University of Sydney. His profound knowledge, rigorous academic approach, rich experience in research, exacting work ethics and easygoing personality have all left a lasting impact on me. His tremendous enthusiasm for work and profound theoretical knowledge will be treasure to me throughout my life. Amiable and respectful, he is the perfect example of a good teacher. He would always take time out of his busy schedule to advise me on my research topic and thesis. Under his influence, I not only gained valuable professional knowledge, but also improved my research methods and even learned to better myself as a person.

My sincerest gratitude also goes to Associate Professor Yonghui Li at the University of Sydney for his elaborate guidance and teachings on the topic of my doctoral studies, which gave me deep insight into the frontier technologies of wireless communications and channel coding. During my stay in Australia, Mr. Li offered me invaluable support and attention not only regarding my academic studies but also for my personal life. If I ever came across any obstacle, he would be there to help me selflessly. Without him, what I have achieved so far would not have been possible. Mr. Li's erudition, uncompromising attitude toward scholarship and scrupulous approach to research will always be an inspiration to me in my future studies and work. I would also like to thank Professor Branka Vucetic at the University of Sydney, whose advice has made this

paper more complete and solid.

During my doctoral studies, I also received help from other mentors and students. I am indebted to their warmth and kindness, which kept me going further like a driving force. I would like to take this opportunity to thank them for their support. Firstly, I want to thank Professor Qing Guo for his help in my study and research. Also, my special thanks go to Associate Professor Zhenyong Wang for his continuous help in my study and life, and to Dr. Min Jia and other teachers for their support. I also want to thank Peng Wang and Jun Li, postdoctoral researchers at the University of Sydney, for their great help in the research and design of my doctoral topic and also for their encouragement and support. I am grateful to all my fellow students at the Telecommunications Laboratory of the University of Sydney. Although I cannot put down each and every of their names here, I want to express my gratitude nonetheless. The days we spent together were joyful and fulfilling, and they will always be a valuable part of my memory. I am grateful to my housemates in Redfern. It is their warm-heartedness and help that made me feel like home and expelled the loneliness in my life of being an international student.

I want to thank the professors and experts reviewing this paper for taking time out of their busy schedule to provide me with valuable comments and suggestions.

I must thank my beloved wife Dan Wu. Without her help, understanding, tolerance and support, I am certain my life in the past few years would have been very different. During my study toward my PhD degree, she took over all the responsibilities for the family. Not only did she have to work to provide extra economic support, she also had to look after our baby daughter Nini. To have her standing behind me gives me the strength to carry on. But for her understanding and hard work, I wouldn’t have come so far.

I also must thank my dear parents, who have put up with lots of hardships and made great sacrifices in order for me to succeed. The utmost care they devoted to my life and the great encouragement they offered regarding my study are invaluable. I hereby wish them a healthy and happy life!

Lastly, my sincerest thanks go to all my friends and those who have offered me help and support.

Statement of Originality

The work presented in this thesis is the result of original research carried out by myself, in collaboration with my supervisors, while enrolled in the School of Electronics and Information Engineering at Harbin Institute of Technology, China and the School of Electrical and Information Engineering at the University of Sydney as a candidate for the Doctor of Philosophy (Cotutelle).

These studies were conducted under the supervision of Prof. Xuemai Gu, Prof. Yonghui Li and Prof. Branka Vucetic. It has not been submitted for any other degree or award in any other university or educational institution.

_____

Bin DUO

School of Electronics and Information Engineering

Harbin Institute of Technology, China

School of Electrical and Information Engineering

The University of Sydney, Australia

March 2016

摘要

随着无线通信技术的快速发展，人们对高速无线数据传输服务的需求也越来越大。但是由于无线频谱资源的日趋紧张，无线通信系统想要支持高速数据传输就必须充分而高效地利用有限的频谱资源，这是未来无线通信技术亟待解决的关键问题。近年来的研究表明，协作通信已经成为解决高速数据传输和大范围网络覆盖的关键技术。协作传输无线数据不仅能够获得空间分集增益，同时还能提高频谱利用率以及系统容量。而中继信道模型作为协作通信网络的重要组成部分更是得到了广泛的研究。分布式信道编码技术是为适应协作通信技术要求而产生的信道编译码技术。好的分布式信道编码方案不仅实现简单，而且可以在充分利用空间分集增益的基础上获得额外的编码增益，还能够进一步提升系统的可靠性能。由此可见，具有性能优异、实现复杂度低的分布式信道编码技术是协作通信系统中需要深入研究的一个核心问题。

基于信道极化现象，Polar 码(Polar Codes)被证明是一种可以达到端到端二元输入离散无记忆信道(Binary-input Discrete Memoryless Channel, B-DMC)对称信道容量的信道编码方案。Polar码作为一种结构化的信道编码技术，具有编码复杂度和译码复杂度都较低，且译码性能不存在错误平层的优异特性。因此，研究协作中继系统中具有低复杂度的 Polar 码编译码方法以及基于 Polar 码的高效协作传输协议以达到中继系统容量具有重要的理论和实用价值。

本文将 Polar 码作为一种强有力的信道编码技术，不仅分析了级联 Polar 码的实际性能，还分别从理论和实际应用两个方面深入对在中继系统中应用Polar码的关键技术进行了研究。论文的主要贡献如下：

第一，针对 Polar 码比特错误率(Bit Error Rate, BER)性能收敛速度较慢以及低密度生成矩阵(Low-Density Generator Matrix, LDGM)码存在高错误平层的问题，本文利用级联编码的思想，提出了一种将 Polar 码作为外码，LDGM 码作为内码的串行级联 Polar-LDGM (Serially-Concatenated Polar-LDGM, SCPL)码方案。首先通过对 Polar 码和 LDGM 码错误平层性能的理论分析，得出了将高码率的 Polar 码作为外码不仅可以保证 SCPL 码性能的收敛速度，同时由于其优异的错误平层特性，通过合理地设计也可使 SCPL 码不存在错误平层的重要结论。然后给出 SCPL 码的编码方法以及基于 Tanner 图的消息迭代译码算法。最后通

过合理地选择 SCPL 码编码设计参数（即内外码速率组合以及内码码重），验证了不同参数下 BER 性能的差异。与低密度奇偶校验(Low-Density Parity-Check, LDPC)码和级联 Polar-LDPC 码相比，外码码长为 N_1、内码码长为 N 的 SCPL 码编码复杂度仅为 $O(N_1\log N_1)+O(N)$。仿真结果表明，SCPL 码具有逼近香农限的性能，并且直到 BER 降到 10^{-10} 也没有出现错误平层。

第二，针对退化半双工单中继信道模型，根据信道极化现象，提出了一种基于 Polar 码的协作部分消息中继转发(Cooperative Partial Message Relaying, CPMR)传输协议。理论分析表明，在该模型中采用 CPMR 方案可以获得渐进达到中继信道容量限的性能，并且推导出在连续取消(Successive Cancellation, SC)译码算法下平均分组错误概率的上界。针对 CPMR 方案在退化半双工单中继系统中的实际应用，首先分析了中继系统容量限，阐释了影响系统容量限的关键参数——时隙分配因子和中继节点与信源节点之间的距离。然后设计了一种基于 Polar 码的联合软信息并行迭代干扰消除(Joint Iterative Soft Parallel Interference Cancellation, JISPIC)接收器。最后对有限码长 CPMR 方案的可行性进行了仿真验证。与单中继系统中传统 LDPC 码方案相比，本文提出的方案继承了 Polar 码所具有的更低编译码复杂度的优点，并且获得了可以与 LDPC 码相媲美的性能。

第三，针对多中继系统中非构造性的随机编码方法编译码复杂度较高的问题，本文将单中继系统中基于低复杂度 Polar 码的 CPMR 传输协议扩展到多中继系统中，中继节点可以根据下一个中继节点或信宿节点需要的可靠消息灵活地转发译码后的部分消息。根据中继节点需要转发的部分消息与部分消息信息比特索引集合之间的对应关系，围绕两种具有正交接收部件的退化多中继网络(Multiple-relay Network with Orthogonal Receiver Components, MRN-ORCs)系统模型，给出了求解相应中继节点待转发部分消息信息比特索引集合的算法，分析了构造性 Polar 码编译码方法，证明了 Polar 码的码长 $N\to\infty$ 且 $\beta\in(0,\frac{1}{2})$ 时，CPMR 方案可以渐进达到这两类退化 MRN-ORCs 的系统容量，同时推导出平均分组错误概率上界仅受限于 $O(2^{-N^{\beta}})$。最后仿真验证了有限码长 CPMR 方案实际应用的可行性。

关键词：协作通信；中继传输系统；分布式信道编码；级联码；Polar 码

Table of Contents

List of Figures

List of Tables

List of Acronyms

AF	Amplify-and-Forward
AWGN	Additive White Gaussian Noise
B-DMC	Binary-Input Discrete Memoryless Channel
BC	Broadcast Channel
BEC	Binary Erasure Channel
BER	Bit Error Rate
BI-AWGN	Binary-Input Additive White Gaussian Noise
BP	Belief Propagation
BPSK	Binary Phase Shift Keying
BSC	Binary Symmetric Channel
CF	Compress-and-Forward
CPMR	Cooperative Partial Message Relaying
DF	Decode-and-Forward
DDMRN-ORCs	Doubly-Degraded Multiple-Relay Network with Orthogonal Receiver Components
DDTRN-ORCs	Doubly-Degraded Two-Relay Network with Orthogonal Receiver Components
EF	Estimate-and-Forward
EXIT	Extrinsic Information Transfer
JISPIC	Joint Iterative Soft Parallel Interference Cancellation
LDGM	Low-Density Generator Matrix
LDPC	Low-Density Parity-Check
LLR	Log-Likelihood Ratio
LR	Likelihood Ratio
MAC	Multiple Access Channel
MAC-GF	Multi Access Channel with Generalized Feedback
MAP	Maximum a posteriori Probability
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MRN-ORCs	Multiple-relay Network with Orthogonal Receiver Components
OTNs	Optical Transport Networks

PEG	Progressive Edge Growth
PF	Parity Forwarding
QF	Quantize-and-Forward
SC	Successive Cancellation
SCLD	Successive-Cancellation List Decoding
SCLDGM	Serially Concatenated Low-Density Generator Matrix
SCPL	Serially Concatenated Polar and Low-Density Generator Matrix
SDMRN-ORCs	Serially-Degraded Multiple-Relay Network with Orthogonal Receiver Components
SDTRN-ORCs	Serially-Degraded Two-Relay Network with Orthogonal Receiver Components
SISO	Soft-Input Soft-Output
SNR	Signal to Noise Ratio
SPIC	Soft Parallel Interference Cancellation
SRN-ORCs	Single-Relay Network with Orthogonal Receiver Components
SU-LLRC	Single-User Log-Likelihood Ratio Calculation
T-MAC	Two-User Multiple Access Channel
TU-MACD	Two-User Multiple Access Channel Detector

Chapter 1 Introduction

1.1 Purpose and Significance of the Research

For the past decades, from the original research on end-to-end wireless communication systems to the present study on multi-terminal wireless network communications, wireless communication technologies have been developing at a rapid pace. At the same time, the demands for wireless communication services are also growing and so are the demands on wireless data transmission rates and communication quality. At present, wireless communications are developing in the direction of achieving bigger bandwidth, higher rate, better reliability and lower realization complexity. To effectively address the fading and interference existing in wireless communications, it has always been a key issue requiring urgent solutions to design communication systems with high reliability, high utilization of spectrum and power, and the ability to increase the channel capacity. As we all know, by using the Multiple-Input Multiple-Output (MIMO) technology [1-3], diversity gain can be obtained through the configuration of multiple antennas at the transmitting terminal. However, due to the limitations imposed by the size and hardware computing power of most wireless terminal equipment, it is usually difficult to apply the MIMO technology directly on such equipment. In order for a single-antenna terminal to also obtain the space diversity gain, Sendonaris et al., employing the concept of cooperative communications [4-10], enabled single-antenna terminals in a wireless network to share each other's antennas through time, frequency or space, which created a virtual MIMO system and eventually obtained the diversity gain (as shown in Fig. 1-1). As a whole new field of research, the cooperative communication technology has prospects

of wide-ranging applications. For example, it can be applied in cellular mobile communication networks, wireless sensor networks, wireless Ad Hoc networks, wireless local area networks and other scenarios. The core idea of cooperative communications is that by taking advantage of the broadcasting properties inherent in a wireless channel, multiple terminals can transmit their own data or help forward data to destinations through different cooperative modes and their corresponding methods of signal processing. Compared with traditional end-to-end wireless communications, the cooperative communication technology not only enables effective sharing of spectrum resources between multiple terminals, but also offers higher data throughput, enhancement of information transmission rates, and improvement in data transmission reliability. Therefore, it has become a central topic in research on next-generation wireless communication systems [11, 12].

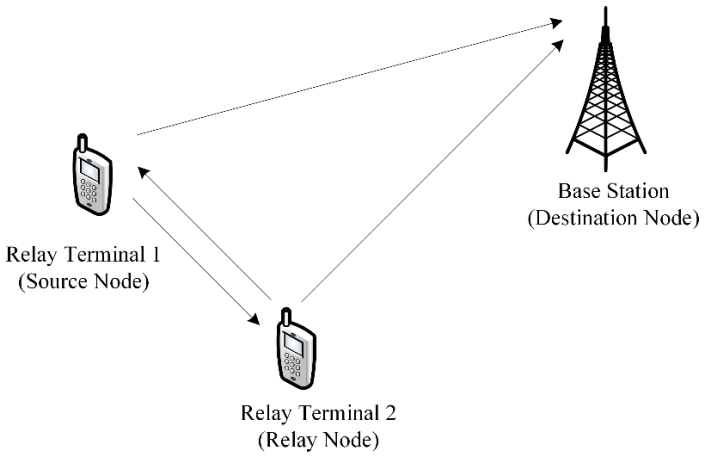


Fig. 1-1 A model of three-terminal cooperative relaying system

To analyze and solve the key issues of the cooperative communication technology, it is firstly essential to establish a controllable mathematical model. The relay channel model as an important component of cooperative communication networks has been extensively studied. In a three-node relay channel, the source node can transmit information reliably to the destination node with the assistance of the relay node [13]. From the perspective of practical applications, it has always been the aspiration of

researchers to design an optimal cooperative communication strategy for relay nodes, and apply it to relay systems. When the relay node and the source node are close in distance, the decode-and-forward (DF) protocol, as a classic cooperative transmission strategy, not only is relatively simple in concept, but also can obtain a near-optimal performance. It can also take advantage of the channel coding technology to play out the strengths of the DF protocol in cooperative communications [14-20].

The distributed channel coding technology is the result of seamlessly combining the cooperative communication technology and the channel coding technology. Based on cooperative communications, by using advanced modern channel codes, the relay node decodes and re-encodes information coming from the source node, after which it cooperates with and complements the coding method of the source node. As a result, it not only achieves additional coding gain on the basis of fully utilizing the cooperative space diversity gain, but also effectively guarantees the reliability of information transmission in cooperative communications. Although the distributed channel coding technology has been a very active area of research in recent years, it still has a lot of theoretical and pragmatic problems which need to be further studied: Firstly, the present cooperative communication model is still one-dimensional and thus it is necessary to investigate the distributed channel coding technology under various channel models. Secondly, it still lacks systematic research results on how to adjust the coding schemes and cooperative transmission strategies applied for the cooperative nodes according to the variation in the network model. Finally, it still calls for an in-depth study on how to design a highly effective distributed channel coding scheme based on a more practical multiple-relay system model.

Despite some modern coding schemes, e.g., the highly mature Turbo codes and Low-Density Parity-Check (LDPC) codes which are able to achieve excellent

performances extremely approaching the Shannon capacity limit, there is still a lack of sufficient yet strict proof for the achievability of the limit. Therefore, it is of great theoretical and practical significance to find a channel coding method, which can be proven to achieve the Shannon capacity limit and has low encoding and decoding complexity simultaneously. The application of distributed Turbo codes and distributed LDPC codes in relay systems has been proven to be able to approach the capacity limit. However, the two codes have high decoding and encoding complexity, respectively. That is why, without a doubt, one key issue to be considered in the practical application of distributed channel coding technologies is regarding how to lower the realization complexity while ensuring an excellent performance. Therefore, to work out a channel coding technology with a simple encoding structure, low decoding complexity and an excellent performance, to apply it to relay systems, and to meet the demands of cooperative communications have together constituted a major motivation to have written this thesis and they have also formed the background of the research.

If this new channel coding method were to be applied in a cooperative communication system, the following challenges need to be met in realizing its distributed coding schemes.

(1) Achievability of capacity

Although distributed Turbo codes and distributed LDPC codes have been verified to have performances capable of approaching the capacity limit of a relay system, they are not yet theoretically perfect as they cannot be proven capable of achieving the capacity limit. Therefore, research on such a channel coding scheme with the ability to reach the capacity limit of a relay system is significant in that it perfects the theory.

(2) Complexity in realization

If a wireless communication network were to achieve high-rate transmission in a

wide area, it should be configured with relatively intensive base stations and repeaters in the network, which would require the wireless terminal equipment to be low in complexity and cheap in price. Distributed Turbo codes and distributed LDPC codes are relatively complex to realize and hence to research on such channel coding technologies of low complexity for wireless networks has value in practical applications.

(3) Design of cooperative transmission protocols

Despite large numbers of research findings on the application of distributed Turbo codes and distributed LDPC codes in relay systems, study on how to utilize the achievability of this channel coding method to design a cooperative transmission protocol suitable for such codes will be of great value as references to relevant research on such codes in relay systems.

(4) Performance in practical applications

In a wireless communication network, how to overcome the channel noise and achieve the maximum transmission rate is a key problem for system design. Therefore, research on the influence of realizing such codes in wireless relay systems on the system performance may serve as valuable references for the application of such codes in a relay system.

1.2 Overview of Channel Coding Theory

In 1948, Shannon published his landmark paper “A Mathematical Theory of Communication”, and ushered in this brand new discipline called information theory, which laid a solid theoretical foundation for the realization of highly efficient and reliable information transmission in communication systems [21]. Shannon pointed out that the basic problem for a communication system to solve is how to efficiently

and reliably transmit information from the source to the destination. However, noise existing in the communication channel will inevitably generate a certain degree of interference on the information being transmitted, and consequently will eventually reduce the reliability of information transmission. Without loss of generality, increasing the redundancy of the information transmitted (this operation is realized by the channel encoder) can make the communication transmission more reliable but it not only decreases the efficiency but also wastes the system transmission bandwidth. Therefore, the key issue for a communication system is the treatment of conflicts between the efficiency and the reliability of a system, i.e., how to overcome the interference in a channel generated by random noise, and reduce the error occurring in the transmission process without lowering the efficiency in information transmission [22-24].

By adopting the random coding technology, Shannon solved the problem regarding the existence of channel codes. This technology is the basic tool for research on information theory. However, due to its high complexity and substantial time delay, completely random block codes do not have any value in practical applications. It was also pointed out by Shannon that there may exist reliable communication at a rate lower than the channel capacity. When the rate is lower than the channel capacity, almost all channel coding schemes will have excellent decoding performance as long as the receiver uses the optimal decoder. However, the complexity of optimal decoding - maximum likelihood (ML) decoding will increase exponentially with the growth of the block length. Hence, due to such difficulties, it is almost impossible to realize optimal decoding in practical applications. That is why researchers in the field of information theory have been committed to designing a channel coding method applicable in practical situations, which can transmit information reliably at a rate

arbitrarily approaching the channel capacity. Their research focuses on the following areas:

- Deriving basic theoretical capacity limits correspondingly based on different communication system models;
- Abstracting different channel coding methods into mathematical models, and using the coding method to support various communication channel models with the expectation of reaching their basic theoretical capacity limits;
- Studying and analyzing the low-complexity algorithms of these channel coding methods.

According to their performances, channel codes can be divided into good codes and bad codes [25]. Good codes can be further divided into very good codes (i.e., codes with arbitrarily small decoding error probabilities when the coding rate is arbitrarily close to the Shannon capacity limit) and sub-good codes (i.e., codes with arbitrarily small decoding error probabilities when the coding rate reaches the non-zero maximum, and is smaller than the theoretical channel capacity). The so-called bad codes refer to the coding methods which will have arbitrarily small decoding error probabilities only when the coding rate decreases to zero. Shannon channel coding theorem shows that although the random codes generated through the random encoding technology are very good codes, they are not practical codes [21].

Since the proposition of Shannon channel coding theorem, many scholars have been dedicated to the study on how to construct a practical very good code capable of approaching Shannon capacity limit. Over time an important branch of information theory came into being, which is the channel coding theory. Through making joint efforts for over sixty years, researchers have found a variety of channel coding methods, which can be applied in practice and can approach the Shannon limit, and

have made significant achievements [26, 27], most of which can be divided into two major categories: algebraic coding and iterative coding. Fig. 1-2 shows the block diagram of the history and evolution of channel coding theory.

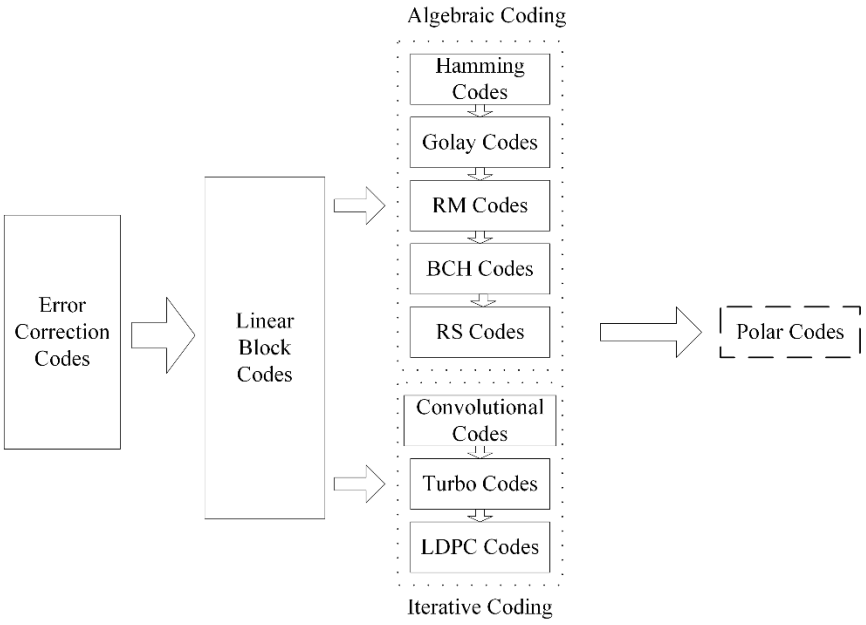


Fig. 1-2 Block diagram of the history and evolution of channel coding theory

1.2.1 Algebraic Codes

Linear block codes have always been a focal point to study in the field of channel coding. Although linear block codes have various algebraic structures, they still have the potential to realize reliable communication at a rate arbitrarily approaching the theoretical channel capacity limit [28-30]. Linear block codes can be expressed through generator matrices, the rows of which constitute the basic vectors of the linear block codes. Meanwhile, linear block codes can also be defined by parity-check matrices, the rows of which constitute the basic vector space orthogonal to such codes. This algebraic structure of linear block codes is beneficial to designing practical encoding and decoding algorithms.

The main method of searching for good block codes is to maximize the minimum distance between their codewords (the minimum distance to a large degree determines the performance of block codes) and at the same time maintain the decoding

complexity at a controllable level. Therefore, the search for good codes was initially regarded as a problem that should be solved through algebraic methods. The best known linear block codes at the early stage include the Golay code [31], Hamming code [32] and Reed-Muller code (i.e. the RM code) [33, 34]. In order to execute decoding, they employ different algorithms to search in code sets for codes which have the minimum Hamming distance to the received codeword sequence (as can be obtained through the hard decision on the observed values received from the channel). By taking advantage of the characteristics of the sub-vector space generated by block codes, such algorithms manage to avoid a brute-force search through all possible codewords, resulting in relatively low complexity. However, the performance of such block codes is quite far away from the Shannon limit. One of the reasons leading to such performances is the method used to search for the minimum Hamming distance, which can only be regarded as the optimal decoding principle for the binary symmetric channel (BSC) and a few others. When information is transmitted in the additive white Gaussian noise (AWGN) channel, the optimal decoding algorithm will be searching in code sets for codewords with the shortest Euclidean distance to the received codewords. On the other hand, to directly make hard decisions on the observed values received from the channel could cause information to be missing, which is another reason for loss in system performances. In addition to the above codes, there are some other very important linear block codes, namely the BCH (Bose-Chaudhuri-Hocquenghem) code, RS (Reed-Solomon) code and General RS (Generalized Reed-Solomon) code [27], [35, 36]. These codes are all composed of elegant but yet complex algebraic structures, so they are also known as algebraic codes in the channel coding field. Many research findings from theoretical studies on algebraic coding played a vital role in the early development of the channel coding theory. Algebraic coding is also an integral part in

practical applications, not only for communication systems, but also for storage systems.

In 1955, Elias invented Convolutional Codes, which was one of the key components of communication systems at that time [37]. Convolutional codes have a somewhat linear structure, and they can be described as a discrete-time finite-state machine. There are several encoding methods for convolutional codes. If encoding is done using the tree structure of convolutional codes, the decoding can be done with the successive decoding algorithm [38, 39]. With the introduction of the threshold decoding algorithm and its successors [40], the decoding performance of convolutional codes has been improved, further advancing its use in applications. Convolutional codes have a good encoding structure property, known as the trellis structure [41], the use of which can facilitate the design of a practical and optimal decoding algorithm [27, 42]. In 1967, Viterbi invented the Viterbi decoding algorithm [43, 44], which has so far already been applied in a range of areas, including channel coding and signal processing. By utilizing linear complexity, the Viterbi algorithm provides a simple method to find the maximum likelihood estimation of transmitted sequences. Meanwhile, as the Viterbi algorithm employs a soft-decision algorithm, resulting in a performance superior to that of a hard-decision decoding algorithm by at least 2dB, this method has been widely recognized as one of the best decoding methods in the AWGN channel. In 1974, Bahl et al. proposed the BCJR decoding algorithm [45], which uses the soft information output from convolutional codes to compute the a posteriori probabilities of the code bits so as to minimize the error probability. Compared with the Viterbi decoding algorithm, the BCJR algorithm has not significantly improved the decoding performance, and it is also very complex. Hence, the Viterbi algorithm has been the standard decoding method of convolutional codes for the following decades. In 1966,

Forney introduced the concept of concatenated codes [46]. Concatenated codes, as a kind of channel coding technology, connect two relatively short codes to form an effective and reliable (relatively long) code. The optimal decoding for concatenated codes adopts an overall coding design, which treats two short codes as one concatenated code, but by merely decoding each code respectively during decoding, a very good performance can be achieved. For the decoding of concatenated codes, the outer codes are usually decoded using hard decisions while the inner codes are usually decoded using soft decisions. Forney's study shows that, while ensuring no significant increase in the complexity of decoding, concatenated codes can also guarantee marked improvement in the decoding performance. Initially, the concatenated coding scheme was only about concatenating two simple block codes but it soon was extended to concatenating convolutional codes. Before the advent of Turbo codes, a serially-concatenated coding scheme, which treats convolutional codes as the inner codes and RS codes as the outer codes, is the best coding method for the AWGN channel [26, 27]. The performance of such concatenated codes is within less than 3dB distance of the Shannon limit of the AWGN channel [47].

1.2.2 Iterative Codes

In 1993, based on the encoding and decoding concepts of convolutional codes and concatenated codes, Berrou et al. proposed a new channel coding scheme, i.e. the Turbo code [48]. As the first instance of a channel coding scheme to demonstrate that reliable communication is achievable in the AWGN channel, the Turbo code boasts the performance within less than 1dB distance of Shannon theoretical capacity limit. Turbo codes are the practical very good codes which academics in the field of channel coding had been longing for; its advent marked the study on channel coding theory entering a groundbreaking era. The basic structure of Turbo codes contains parallel

concatenated convolutional encoders, where one convolutional code is the interleaving form of another. Codewords generated through this constructing method will to a certain degree be both structural and pseudo-random. The Turbo decoding algorithm is iterative soft information decoding based on every convolutional code. Each component decoder decodes based on the BJCR algorithm. During the iterative process, such component decoders exchange their soft output information, and utilize the output probabilities of their counterparts as their own a priori information. The fundamental reason why the Turbo code enjoys a performance approaching the Shannon channel capacity limit is that it realizes the idea of random coding put forward in Shannon theory through the pseudo-random permutation of information sequences by a random interleaver, and obtains a kind of pseudo-random long codes. This lays the foundation for the application of Shannon random coding theory. In the meantime, due to the use of an iterative decoding structure, the decoding performance is improved through the iterative exchanging and constant revising of soft information transmitted between two component decoders [49, 50]. Such an iterative decoding method of Turbo codes is also called the “Turbo principle”. At present, the idea embodied in the “Turbo principle” has permeated almost all the fields of advanced signal processing technologies for physical layers. It dramatically drives the progress and development of communication technologies through the combination with technologies such as equalization, modulation, space-time coding, orthogonal frequency division multiplexing, and so on.

Just as the advent of the Turbo code and the introduction of the ideas on iterative decoding opened a new chapter for modern coding theory, the LDPC code [28, 29] rediscovered by Mackay et al. established a new milestone for the modern development of channel coding theory [51]. LDPC codes are a type of linear block codes with sparse matrix structures, which not only can directly apply Shannon’s idea

of constructing random linear block codes of sufficient block lengths, but also can use iterative decoding algorithms of low complexity. Due to the use of the belief propagation (BP) algorithm, the decoding performance of LDPC codes with linear decoding complexity can approach the Shannon channel capacity limit [52-55]. For medium to long block lengths, the performance of LDPC codes has already surpassed that of Turbo codes. Also LDPC codes enjoy lower decoding complexity and can simultaneously carry out parallel decoding and detect decoding errors. Therefore, LDPC codes have become a central research topic in the field of channel coding. Up until now, these codes have been regarded as very good codes for their performance in applications [49, 50]. LDGM codes, as a special form of LDPC codes, have attracted attention for their relatively low complexity. LDGM codes, which are linear systematic codes, boast much lower encoding complexity than that of LDPC codes and at the same time, their parity-check matrices still have some level of sparsity, which is why the decoding of LDGM codes can also employ the BP decoding algorithm [56]. However, as there exist identity matrices in the parity-check matrices of LDGM codes, there are conspicuous error floor problems when decoding LDGM codes [51]. In order to effectively lower the error floors, Ref. [56] proposed a serially-concatenated LDGM coding scheme, which not only maintains relatively low decoding complexity of LDGM codes, but also obtains an excellent performance approaching the Shannon channel capacity limit [56]. Hence, in recent years, much attention has been devoted to research on serially-concatenated LDGM codes [57-60].

Turbo codes and LDPC codes represent two kinds of channel coding schemes approaching the Shannon limit in the field of channel coding. As long as the block length is ensured to be sufficiently long, the performances of such codes can closely approach the Shannon channel capacity limit. However, in practical applications, the

complexity in encoding and decoding is a key issue that must be taken into consideration. In general, although the encoding complexity of Turbo codes is relatively low, but their decoding complexity is relatively high [61]. In comparison, LDPC codes have relatively low decoding complexity but relatively high encoding complexity [62]. Therefore, to identify a channel coding scheme which integrates both good performance and low complexity has become a critical research topic.

During the past decades, the development of algebraic coding and iterative coding has made significant progress, especially in the in-depth study on modern encoding and decoding methods with excellent performance, such as Turbo codes and LDPC codes. Latest achievements in this field have extremely approached the final goal of the AWGN channel - the Shannon limit. However, it still remains an ultimate target for researchers to succeed in identifying practical very good codes with the following characteristics:

- Capable of transmitting information at a rate that arbitrarily approaches the channel capacity;
- Enjoying low encoding and decoding complexity;
- Guaranteeing reliable communication through such codes, which can be theoretically proven.

1.2.3 Polar Codes

In 2009, based on the phenomenon of channel polarization, Arikan proposed a brand-new type of channel coding technology - the polar code [63]. Polar codes combine the encoding concept underlying algebraic coding and the decoding concept underlying iterative coding, and meet all the requirements of practical very good codes as mentioned above:

- Arikan proved that polar codes can achieve the symmetric capacity of binary-

input discrete memoryless channels (B-DMC), where symmetric channels include the binary erasure channel (BEC), the BSC channel and the binary-input additive white Gaussian noise (BI-AWGN) channel.

- As a coding method with low encoding and decoding complexity, polar codes have prospects for being put into practical application. Arikan proved that when the block length is N, the time and space complexities of the encoding and decoding algorithms are both $O(N \log N)$, where, $O(N)$ is the big O notation (Landau notation), describing the asymptotic performance of the function.
- Arikan inferred that, under the successive cancellation (SC) algorithm, the average block error probability of polar codes is upper bounded by $O(2^{-N^{\beta}})$ [64], where $0 < \beta < 1/2$.

Fig. 1-3 depicts the system model of channel polarization. The polarization of a B-DMC results in two types of polarized bit channels: the "good" bit channel (i.e., binary channel with channel capacity close to 1 bit/channel use) and the "bad" bit channel (i.e., binary channel with channel capacity close to 0). The capacity of those "good" bit channels equals the mutual information with equal probability input (that is, equal to the symmetric capacity of discrete memoryless channels). The encoding of polar codes is to let the information bits to be encoded pass through the bit channels approaching the channel capacity (here channel polarization can be regarded as one kind of precoding or coding preprocessing). At the same time, those frozen and pre-determined bits are transmitted to bit channels whose channel capacity is close to 0. Because the predetermined bits are the key component of the SC decoding process for polar codes, these frozen and pre-determined bits cannot be ignored. Therefore, information bits to be encoded and the frozen bits make up the input bit sequence $(U_1, U_2, \ldots, U_N)$, which, through channel polarization transformation, produces the

coding sequence $(X_1, X_2, \ldots, X_N)$. After being passed through the channel, the received sequence $(Y_1, Y_2, \ldots, Y_N)$ is decoded by the receiver to get the information bits. Additionally, Korada also pointed out that in a physically degraded model, those bit channels which are "good" for degraded channels must be "good" bit channels for non-degraded channels as well [65].

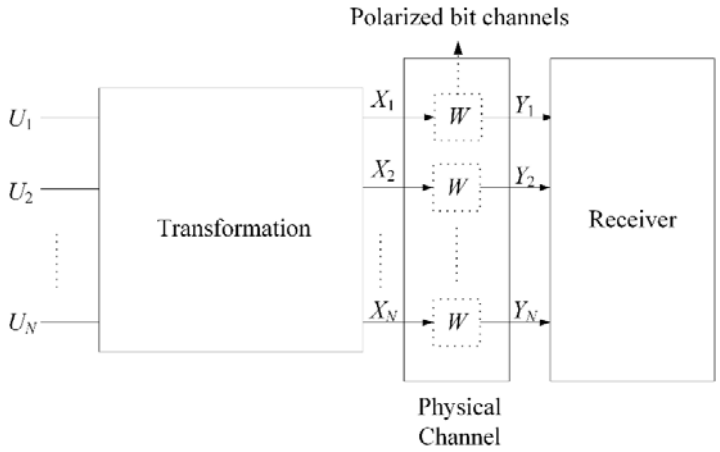


Fig. 1-3 The theoretical model of channel polarization for polar codes

Presently, the theoretical study and practical application of polar codes are still in the early stage, but channel polarization and related technologies have been used to solve some important issues in relation to information theory. For example: Arikan et al. extended the channel polarization in the B-DMCs to arbitrary discrete memoryless channels, and proved that similar to the case of the B-DMCs, polar codes can also achieve the capacity of arbitrary discrete memoryless channels [66]. Hassani [67] and Tanaka et al. [68, 69] further investigated the rate of channel polarization and derived the upper bound on the optimal achievable block error probability of polar codes under the SC decoding algorithm. Based on this research finding, Korada et al. analyzed and summarized the general construction standards for polar codes and the algorithms for the bounds on channel polarization exponents [65, 70]. Tal and Vardy [71, 72] respectively studied the encoding and decoding methods for polar codes. The findings

show that the polar codes constructed by Tal and Vardy only have time complexity of $O(N)$ and space complexity of $O(\log N)$. Meanwhile, the successive-cancellation list decoding (SCLD) algorithm proposed by Tal and Vardy, improved the original SC decoding algorithm, and achieved an excellent performance comparable to that of LDPC codes. Hassani et al., who applied the polar codes to compound channels, proved that in general, under the SC decoding algorithm, the compound channel capacity based on polar codes is strictly less than the capacity of the compound channel [73]. Goela et al. analyzed and constructed a polar coding scheme for the deterministic broadcast channel (DBC). Their research results show that the proposed scheme not only inherits the block error probability of end-to-end polar codes, but also maintains the decoding complexity of $O(N \log N)$ [74]. Regarding the dual channel of the DBC channel, i.e., the multiple access channel (MAC), research conducted by Abbeand Telatar et al. [75, 76] shows that the channel polarization phenomenon still exists when polar codes are applied in the arbitrary multi-user MAC channel. They not only designed a polar code construction scheme in the MAC channel, but also proved that the constructed polar codes maintain the properties of end-to-end polar codes (i.e., complexity and the upper bound on the block error probability). Such a scheme serves as important theoretical basis for research undertaken in this thesis.

1.3 Research Status of Distributed Channel Coding

Research on the basic theory of distributed channel coding includes the capacity of relay systems and corresponding channel coding methods.

1.3.1 Capacities of Relay Systems

If a source node transmits information to a destination node without the assistance of any other communication terminals, such communication is referred to as direct

communication, single-user communication or end-to-end communication. For cooperative communications, there is at least one extra communication terminal in the communication network, which can help forward the source information. Meulen brought in the classic three-terminal relay channel [13], and derived the upper and lower bounds on the relay channel capacity. Although up until now the closed-loop expression of the capacity of general relay channels remains unknown, Cover and Gamal [77] significantly improved the upper and lower bounds derived by Meulen through extensive research. Without a doubt, Cover and Gamal's findings played a decisive role in the development of relay channel capacity. Many of their important conclusions still remain irreplaceable so far. Cover and Gamal designed a classic relay cooperative transmission strategy - the DF protocol [11], [78, 79], and proved that by using this protocol, the achievable rate of relay systems can be attained [77]. There are mainly three different encoding and decoding methods for the DF protocol, which are:

- Method 1: irregular encoding/successive decoding;
- Method 2: regular encoding/sliding-window decoding;
- Method 3: regular encoding/backward decoding.

With Method 1, Cover and Gamal separately used the block Markov superposition encoding, random binning and successive decoding, and used codebooks with different block lengths. Therefore this method is called irregular encoding. King [80] researched on the multi-access channel with generalized feedback (MAC-GF), derived the achievable rate region of the channel, and summarized the research findings of Slepian and Wolf [81], Gaarder and Wolf [82], and Cover and Leung [83]. Utilizing the encoding and decoding methods of Method 2 (i.e., the source and the relay codebook share the same length, and the destination employs the sliding-window decoding algorithm), Carleial [84] further expanded King's research results, not only deriving

17 bounds regarding the achievable rate region, but also obtaining the same achievable rate as that of Cover and Gamal's through properly choosing some random variables. Based on the MAC-GF model, Willems [85] introduced the backward decoding technology, i.e., Method 3, which even outperforms the sliding-window decoding algorithm [86-88]. He also pointed out that, for relay channels, using Method 3 can obtain the same achievable rate as can be obtained by using Method 1 and Method 2.

Extensive research has also been carried out on the expansion of the above DF single-relay cooperative strategies into multiple-relay systems, and the design of corresponding encoding and decoding methods. Aref [89] was the first to have applied Method 1 in degraded relay networks [90], and he designed "binning" strategies capable of achieving the capacity of deterministic broadcast relay networks and deterministic interference-free relay networks. The achievability of capacity was proved using the concept of cut-set theoretical limits [77, 89], which have become a standard tool in defining channel capacity regions [91]. Similar to what Aref did, Gupta and Kumar [92] applied Method 1 to DF multiple-relay networks. What the two methods have in common is that they both consider the relaying process as a multi-hop strategy, which means that before the source information arrives at the destination node, the relay node has already performed successive decoding on the received source information. Following that, Xie and Kumar [93, 94] designed Method 2, the encoding and decoding methods for DF multiple-relay networks, and demonstrated that their method can obtain an achievable rate higher than Aref's and Gupta's. Kramer et al. [95] further probed into Method 3, and pointed out that the two types of coding strategies have the same achievable rate, but sliding-window decoding has shorter coding delay than backward decoding, so Method 2 became the then most simple strategy which could reach the optimal achievable rate. The above multi-hop strategies

require all the relay nodes to decode the received source messages, so the source rate is limited by the decoding abilities of those relatively poor source-relay links. If the relay nodes can flexibly select appropriate message subsets to decode (these decoding sets not only include source messages, but may also include messages forwarded from other relay nodes), then multi-hop DF rates will be further increased. Stemming from this idea, Razaghi and Yu [96] proposed a new DF cooperative strategy - the parity forwarding (PF) protocol, based on which they also designed the irregular encoding/joint decoding method, referred to as Method 4 in this paper. By using this method, Razaghi and Yu proved that the PF protocol is capable of achieving the capacity of degraded relay networks.

As we can see, so far research on multiple-relay systems has been fruitful, but most of the work reached their main conclusions based on non-constructive random coding methods. These conclusions indicate that there are channel codes capable of achieving the capacity of relay systems, but how to design practical coding schemes with low encoding and decoding complexity as well as the ability of achieving the capacity remains a challenge.

1.3.2 Distributed Channel Coding Technologies for Relay Systems

As an important breakthrough for modern wireless communications [10], cooperative communication technologies, taking advantage of the broadcast properties of wireless channels, establish spatial parallel communication channels, enable multiple single-antenna terminals to obtain space diversity gain through sharing, and realize virtual MIMO systems, which can improve wireless communication reliability, data transmission rates, utilization of spectrum resources and system capacity without increasing the transmitting power and system bandwidth. The application of distributed channel coding technologies brought further development to cooperative

communications, and based on cooperative diversity, distributed channel coding technologies further enhanced the communication performance of systems.

With the development of network information theory, distributed channel coding technologies of multifarious types sprang up in an endless fashion while relevant research on the application schemes for the classic relay channel model has also been active and dynamic. Hunter and Nosratinia [97] firstly proposed the concept of "coding cooperation" based on convolutional codes, and then this concept was extended into space-time coding [98, 99]. However, these distributed coding schemes are still quite far away from the capacity limit of the relay channel, which triggered research on applying Turbo codes to relay channels [15], [100, 101]. All these distributed Turbo coding schemes have considered an orthogonal receiving channel, which simplifies the structure of the receiver at the expense of spectrum effectiveness. In order to achieve higher system capacity, Zhang and Duman [16, 17] separately proposed distributed Turbo encoding and decoding methods for the full-duplex and the time-division half-duplex relay channel models, and they, through analyzing destination nodes, properly designed a MAC channel detector and an iterative decoding algorithm. Their proposed scheme is capable of approaching the capacity limit of the relay channel. Meanwhile, in order to effectively address the unwanted impact on systems caused by relay decoding errors, Li et al. [102] proposed a distributed Turbo coding and decoding scheme based on soft information relay cooperation. As LDPC codes are very good codes extremely approaching the Shannon limit, research on applying LDPC codes in relay channels has attracted wide attention. Studies on various distributed LDPC coding schemes have led to excellent performances approaching the capacity limit in the full-duplex and/or time-division half-duplex relay channel model [18-20], [103-106]. As previously mentioned, polar codes have the advantages of low

complexity and being capable of achieving the capacities of B-DMCs, which Turbo codes and LDPC codes lack. Therefore, it has been a new direction to apply polar codes in wireless relay networks, and explore distributed polar encoding and decoding schemes which meet the needs of cooperative communications. Andersson et al. [107] were the first to have applied polar codes in the DF single-relay channel, and proved that polar codes with a nested structure can achieve the capacity of binary-input physically degraded relay channel with orthogonal receiver components. Then, Serrano et al. [108] extended Andersson's findings into relay channels with arbitrary inputs, and proved the achievability of the capacity. Through the implementation of block Markov coding, it was proven that polar codes can achieve the capacity of general physically degraded relay channels (i.e., without assuming orthogonal reception) [109]. All the above work was based on the full-duplex mode (i.e., relay nodes can send and receive data at the same time or within the same frequency band) and/or the assumption that destination nodes have orthogonal receiving components. Such an assumption bypasses the need to consider interference existing in received signals at the expense of spectrum width. In practical applications, in order to simplify the system design, people tend to make relay nodes work in half-duplex mode, and in order to make full use of system resources, the source and relay nodes often transmit signals within the same frequency band. So far, there has been no research regarding the application of polar codes in actual half-duplex relay systems and multiple-relay systems.

1.4 Main Research Contents of This Doctoral Thesis

This research topic originates from “Cooperative Communications for Future Wireless Networks”, a doctoral training program co-supported by Harbin Institute of

Technology and the University of Sydney. Because of the natural properties of distributed broadcasting innate in wireless communication networks, cooperative and distributed signal processing technologies can be used to markedly reduce system power consumption and enhance system performance. The relay protocol, distributed channel coding and cooperative signal processing are key technologies for designing cooperative wireless networks, where there are still many research topics demanding immediate solutions. The main objective of this project is to research on relevant cooperative communication technologies for multiple-relay networks, including optimization of distributed channel coding, design of joint channel and network coding, study on new relay protocols, configuration of cooperative transmission schemes, allocation of resources and power, and design of cross-layer optimization.

Polar codes are among the first category of channel coding methods to have been proven capable of approaching the capacities of B-DMCs. The introduction of the polar code is of great theoretical significance. That is to say, not only did its advent verify the existence of channel coding schemes approaching Shannon channel capacity limit, but also it enjoys relatively low encoding and decoding complexity. However, despite the perfect theoretical proof of the polar code, in practice, the decoding performance of polar codes with finite block lengths needs to be further improved [110, 111]. In order for polar codes to acquire better performances in practical applications, much literature has been written on the study of polar decoding algorithms, and notable progress has been made in improving the performance of polar codes [71, 72], [112-121]. Therefore it is of great practical importance to utilize the advantages of polar codes, and apply them in real-world communication environment, which can transform polar codes eventually into practical very good codes. With the polar code as its focus, this thesis firstly, based on the analysis of advantages and disadvantages

of polar codes and LDGM codes, discusses the feasibility of serially-concatenated polar-LDGM (SCPL) codes and studies the SCPL coding scheme. Then building upon the analysis of the relay systems and the corresponding channel capacities, this thesis proposes a new DF relay cooperative transmission strategy, which this thesis calls the cooperative partial message relaying (CPMR) protocol, and studies practical decoding methods for polar codes under such a protocol.

This thesis is composed of five chapters. Apart from this chapter, which serves as Introduction, the contents of the remaining chapters are arranged as follows:

Chapter 2: Channel Polarization and Polar Codes

This chapter, based on Shannon information theory, presents the basic definitions of channel capacity and the Shannon limit, which are of relevance to channel coding technologies. Then it analyzes the phenomenon of channel polarization in details, with emphasis on the coding methods for polar codes, summarizes the advantages and disadvantages of polar codes in theory and in practical applications, and fully quantifies the encoding and decoding complexity of polar codes.

Chapter 3: Encoding and Decoding Algorithms for SCPL codes

This chapter mainly investigates the methods for realizing the SCPL coding and their performance. Based on the analysis of the reasons why LDGM codes generate high error floors, the chapter, incorporating polar codes' advantage of no error floors and utilizing the idea of concatenated codes, proposes a SCPL coding scheme, which uses polar codes with high coding rates as the outer codes and LDGM codes as the inner codes. Then, the chapter studies the concatenated coding method for SCPL codes and derives the message iterative decoding algorithm based on the Tanner graph. By properly selecting design parameters for SCPL coding, its performance is evaluated by simulations. The performance and encoding complexity of SCPL codes are

compared with those of LDPC codes and the serially-concatenated Polar-LDPC coding scheme under the same conditions.

Chapter 4: Polar Coding Schemes for Single-Relay Transmission Systems

This chapter mainly discusses the encoding and decoding methods for polar codes in degraded half-duplex relay systems. Firstly, the chapter analyzes the channel capacity of degraded half duplex single relay channels, based on which, incorporating the coding structure of polar codes, the chapter proposes a CPMR protocol with low complexity, and designs the corresponding encoding and decoding methods for polar codes. The findings of theoretical research show that to apply a polar coding scheme based on the CPMR protocol in such a system can obtain a performance asymptotically approaching the capacity of the relay channel. The study also derives the upper bound on block error probability under the SC decoding algorithm. Then, to verify the feasibility of the CPMR protocol, an actual system model is built and a joint iterative soft parallel interference cancellation (JISPIC) receiver structure based on polar codes is designed. Through the setting up of a simulation platform, some key parameters impacting the channel capacity limit are analyzed, such as time allocation factor, power allocation and the distance between relay and source. Finally, the CPMR scheme with finite block lengths is compared and analyzed through simulations.

Chapter 5: Polar Coding Schemes for Multiple-Relay Transmission Systems

In response to the relatively high encoding and decoding complexity of the non-constructive random coding method for multiple-relay systems, this chapter is mainly concerned with extending the CPMR protocol with low complexity proposed in Chapter 4 into multiple-relay networks, and designing encoding and decoding methods for polar codes based on the characteristics of the system model. Firstly, the system capacity of the general multiple-relay networks is given, based on which, the system

capacities of two types of multiple-relay network with orthogonal receiver components (MRN-ORCs) are derived. Then, regarding these two types of system model, the chapter respectively describes the correlations between the partial messages and the information sets of the partial messages for corresponding CPMR protocols, analyzes the constructive encoding and decoding process of polar codes, derives the upper bound on the average block error probability under the SC encoding algorithm, and proves that in these two systems, the CPMR protocols based on polar codes can completely replace the traditional DF protocol based on highly complex random encoding methods and can approach the system capacity with arbitrarily small error probabilities. Lastly, the feasibility of applying the CPMR schemes to multiple-relay systems is verified through simulations.

Chapter 2 Channel Polarization and Polar Codes

2.1 Introduction

Shannon channel coding theorem proved the existence of a coding method capable of achieving the channel capacity, but it did not provide any specific code construction method. Although the theorem has been around for more than half a century, it was not until the last twenty years that some modern channel coding methods have been verified to be able to approach or even proven to be capable of achieving the Shannon limit. Based on Shannon information theory, this chapter begins by presenting the basic concepts related to channel coding technologies, such as information entropy, etc. Then on the basis of mutual information, the chapter gives the important definitions of channel capacity and the Shannon limit, with emphasis on channel capacity limit, which is the final goal to be achieved for channel coding technologies in communication systems. Then the chapter studies in detail the basic principles of channel polarization and the encoding and the decoding algorithms of polar codes, derives the encoding and decoding complexity of polar codes and analyzes the advantages and disadvantages of polar codes in theory and in practical applications. The invention of polar codes serves as strong evidence for the existence of a coding method capable of achieving the channel capacity proposed in Shannon channel coding theorem. Polar codes not only can achieve the symmetric capacities of the B-DMCs, but also boast relatively low encoding and decoding complexity. Therefore, the study on polar codes has great theoretical and practical significance.

2.2 Information Entropy

Information entropy is a measure of information [21]. For an arbitrary probability distribution, the entropy measures the uncertainty of random distribution of such information.

The entropy $H(X)$ of a discrete random variable X is defined as

$$H(X) = E[-\log p(X)] = -\sum_{x \in \mathcal{X}} p(x) \log p(x), \tag{2-1}$$

where $E[.]$ evaluates the mathematical expectations of random variables, $\mathcal{X}$ is the alphabets where X takes value for $x \in \mathcal{X}$ and $p(x)$ denotes the probability mass function of X.

The unit of entropy is related to the numeric value of the logarithmic base in Equation (2-1). Most commonly, the logarithmic base is 2 and the unit of entropy is bit. For theoretical derivation, e is often used as the base while nat is used as the unit. Other bases and units can also be adopted, and they are all interconvertible.

Similarly, according to the joint probability distribution and conditional probability distribution of multiple random variables, the definition of entropy can be extended to the joint entropy and the conditional entropy of multiple random variables.

If the joint probability mass function of binary discrete random variables (X, Y) is $p(x, y)$, the joint entropy $H(X, Y)$ can be defined as

$$H(X, Y) = E[-\log p(X, Y)] = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y). \tag{2-2}$$

If the joint probability mass function of binary discrete random variables (X, Y) is $p(x, y)$, the conditional entropy $H(Y \mid X)$ can be defined as

$$H(Y \mid X) = E[-\log p(Y \mid X)] = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y \mid x). \tag{2-3}$$

According to the definition of conditional entropy, for two random variables X and

Y, their conditional entropy $H(Y \mid X)$ denotes the remaining uncertainty of Y when the entropy of X is known, and the entropy with decreased uncertainty is the mutual information.

If the joint probability mass function of binary discrete random variables (X, Y) is $p(x, y)$, and theirs marginal probability mass functions are $p(x)$ and $p(y)$ respectively, then the mutual information $I(X;Y)$ is

$$I(X;Y) = H(X) - H(X \mid Y)$$

$$= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \tag{2-4}$$

Based on the definitions of joint entropy, conditional entropy and mutual information of binary random variables, the chain rule of mutual information of multiple random variable can be deduced as follows: The joint probability mass function of random variable $X_1, X_2, \cdots, X_N$ is $p(x_1, x_2, \ldots, x_N)$, and hence [91]

$$I(X_1, X_2, \ldots, X_N; Y) = \sum_{i=1}^{N} I(X_i; Y \mid X_{i-1}, X_{i-2}, \ldots, X_1). \tag{2-5}$$

Similarly, the entropy, joint entropy, conditional entropy and mutual information of continuous random variables can also be defined.

2.3 Channel Capacity and the Shannon Limit

Speaking in general terms, channels are signal paths with transmission medium as the basis, such as telephone lines, air, and the Internet. For communication systems, the main role of channels is to transmit information. Fig. 2-1 shows the diagram of a basic end-to-end communication system. Usually two symbols are used to represent the input and output of a channel while the characteristics of a channel are usually denoted by transition probabilities. That is to say, if a particular signal is used as a

channel input, each pair of input and output signals, together with the transition probabilities, could characterize the physical channel. To characterize a channel with this probability mapping not only simplifies the channel model but also describes its most important physical properties.

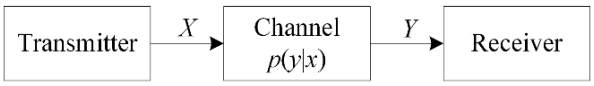


Fig. 2-1 Block diagram of a communication system

A discrete channel is a system composed of input alphabets $\mathcal{X}$, output alphabets $\mathcal{Y}$ and a transition probability matrix $p(y|x)$, where $p(y|x)$ is the probability of receiving an output symbol *y* when a transmission symbol *x* is known. If the probability distribution of an output is only dependent on its corresponding input, and is independent of the channel's previous input and output, the channel is called a memoryless channel.

A discrete memoryless channel's channel capacity *C* is

$$C = \max_{p(x)} I(X;Y), \tag{2-6}$$

which has the following characteristics [21]: For an arbitrary given $\epsilon > 0$, if the information transmission rate $R \le C$, there must exist a channel coding method, which makes the maximum decoding error probability arbitrarily less than or equal to ϵ with the increase of block lengths. The proof of the existence of such a channel coding method should meet the following three conditions:

1) a random coding construction method is adopted;

2) the block length is sufficiently long;

3) an optimum ML decoding algorithm is used for decoding.

Fulfilling the above three necessary conditions, Shannon proved in theory that error-free transmission with rates arbitrarily approaching the channel capacity is achievable.

On the contrary, if the information transmission rate $R > C$, there would not be such a channel coding method to ensure the reliability of information transmission.

The above discussion indicates that channel capacity is the maximized mutual information. Channel capacity can also be defined as the maximum rate at which a channel transmits information with an arbitrarily low error probability, with the unit being bit/channel use. To take one step further, the situation with discrete channels can also be extended to that with continuous random channels.

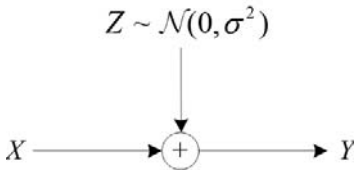


Fig. 2-2 Diagram of a Gaussian channel

Fig. 2-2 shows the model of the AWGN continuous channel. The channel input is a continuous signal *X*, the output is a continuous signal *Y* and the noise is *Z*, where *Z* is the Gaussian distribution subject to variance $\sigma^2 = N_0/2$, i.e. $Y = X + Z$, $Z \sim \mathcal{N}(0,\sigma^2)$, and hence there is $H(Z) = \log(\sqrt{2\pi e}\sigma)$. Assume that *Z* and *X* are independent of each other. When the input average power is limited by *P* and the continuous entropy of normal distribution is maximum [91], from this we can get

$$H(Y) \le \log(\sqrt{2\pi e(P+\sigma^2)}). \tag{2-7}$$

Therefore, the capacity of the Gaussian channel is

$$\begin{aligned} C &= \max_{p(x):E(X^2)\le P} I(X;Y) \\ &= \max_{p(x):E(X^2)\le P} (H(Y) - H(Z)) \\ &\le \log(\sqrt{2\pi e(P+\sigma^2)}) - \log(\sqrt{2\pi e}\sigma) \\ &= \frac{1}{2}\log_2(1+\frac{P}{\sigma^2}), \end{aligned} \tag{2-8}$$

where the equality holds when $X \sim \mathcal{N}(0, P)$.

In the domain of channel coding theory, E_b / N_0 is often the preferred method to characterize channel capacity, where E_b is the average energy of each information bit (information bit before encoding). Define signal-to-noise Ratio (SNR) $\mathrm{SNR} = E_b / N_0$. Suppose a coding rate is given as *R*, and each symbol delivers 2 bits, then

$$E_b = \frac{E_s}{2R}, \tag{2-9}$$

where E_s denotes the average energy of transmitted symbol (modulation symbol after encoding).

Then, Equation (2-8) can be further rewritten as

$$C = \frac{1}{2}\log_2(1 + 2R\frac{E_b}{N_0}). \tag{2-10}$$

In Equation (2-10), the upper bound on the coding rate *R* is given by the channel capacity *C* of the AWGN channel with continuous input and continuous output, i.e. $R \le C$. Therefore, the coding rate *R* and E_b / N_0 have the following relations:

$$R \le \frac{1}{2}\log_2(1 + 2R\frac{E_b}{N_0}) \tag{2-11}$$

$$\Rightarrow \frac{E_b}{N_0} \ge \frac{2^{2R} - 1}{2R} \tag{2-12}$$

$$\Rightarrow \frac{E_b}{N_0} \ge \lim_{R \to 0} \frac{2^{2R} - 1}{2R} = \ln 2 \approx -1.59\mathrm{dB}. \tag{2-13}$$

Equation (2-13) indicates, with a given coding rate *R*, the minimum SNR threshold for realizing error-free transmission in an AWGN channel with continuous input and continuous output, i.e., the Shannon limit. With the increase of the coding rate *R*, the Shannon limit also increases. As *R* approaches 0, the Shannon limit can obtain the minimum value.

If the input X of the above AWGN channel is a binary discrete random variable with equal probability distribution, the AWGN channel is called a BI-AWGN channel. If X takes value from alphabets $\{a,-a\}$, meaning that a binary phase-shift keying (BPSK) of amplitude a is used for modulation, then it satisfies [122]

$$I(X;Y) = -\int_{-\infty}^{\infty} \phi(y,E_b,\sigma^2)\log_2 \phi(y,E_b,\sigma^2)dy - \frac{1}{2}\log_2 2\pi e\sigma^2, \qquad (2\text{-}14)$$

where $\phi(.)$ is defined as

$$\phi(y,a,\sigma^2) = \frac{1}{\sqrt{8\pi\sigma^2}}\left[e^{-(y-a)^2/2\sigma^2} + e^{-(y+a)^2/2\sigma^2}\right]. \qquad (2\text{-}15)$$

Fig. 2-3 compares the curves of channel capacities of the AWGN channel and BI-AWGN channel versus SNR. Take the AWGN channel for example. When SNR is 1dB, as derived from the conclusion of Shannon coding theorem, as long as the information transmission rate is lower than 0.5 bit/channel use, a coding method can be found to realize error-free information transmission in the AWGN channel. On the other hand, when the information transmission rate is 0.5 bit/channel use, as long as $\text{SNR} < 1\,\text{dB}$, error-free information transmission can be guaranteed. However, when $\text{SNR} > 1\,\text{dB}$, provided that channel coding methods are not considered, error-free information transmission cannot be guaranteed. It can also be seen from Fig. 2-3 that, as SNR increases, the channel capacity of the AWGN channel will gradually exceed 1 bit/channel use, and the maximum channel capacity of the BI-AWGN channel is 1 bit/channel use. That is to say, when binary information is input into an AWGN channel, the channel can only obtain 1 bit of useful information. Therefore, there could always be a relation between the AWGN channel capacity and the BI-AWGN channel capacity as follows:

$$C_{\text{AWGN}} > C_{\text{BI-AWGN}}. \qquad (2\text{-}16)$$

However, at a very low SNR, C_{AWGN} and $C_{\text{BI-AWGN}}$ are almost equal.

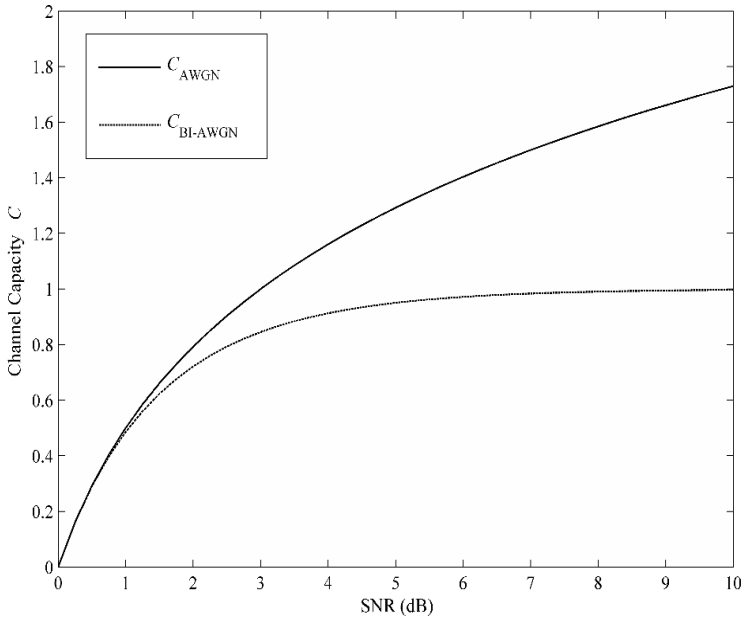


Fig. 2-3 Capacities of the AWGN channel and the BI-AWGN channel

As pointed out by Shannon channel coding theorem, if the information transmission rate is smaller than the channel capacity, information can be transmitted without error. Based on this theorem, making full use of separation theorem and rate distortion theory, Cover studied methods to determine the maximum information transmission rate under conditions where error-free transmission cannot be guaranteed, or to put it differently, methods to compute the minimum SNR for reaching a transmission rate when such a transmission rate and the error probability are given [91, 122].

Suppose the coding rate is R and the error probability of information transmission is p_b. Then based on rate distortion theory, we can know that the quantity of information carried by each transmission symbol is $R(1-H(p_b))$. To make sure that information is transmitted from source with arbitrarily small error probabilities, $R(1-H(p))$ should be smaller than the channel capacity C. Therefore, the maximum achievable rate $C^{(p)}$ with average distortion (i.e., bit error probability) p_b can be expressed as

$$C^{(p)} = \frac{C}{1 - H(p)}. \tag{2-17}$$

Fig. 2-4 demonstrates the SNR E_b / N_0 required to transmit information at various transmission rates in the AWGN channel and BI-AWGN channel. For any curve made in this figure, the area to its right is achievable. Thus, these curves are the ultimate goals to be achieved by the design of very good codes, meaning that when the performance curve of such very good codes coincides with a curve in this figure, we can say that such very good codes transmit information at a rate equal to the channel capacity. For Fig. 2-4, attention needs to be paid to the following:

1) when SNR is very low, the performance of the BI-AWGN channel and that of the AWGN channel are quite similar, which is in line with the curves shown in Fig. 2-3;

2) a higher coding rate demands higher SNR;

3) the vertical asymptotic line (i.e., BER = 0) is precisely the channel capacity *C*.

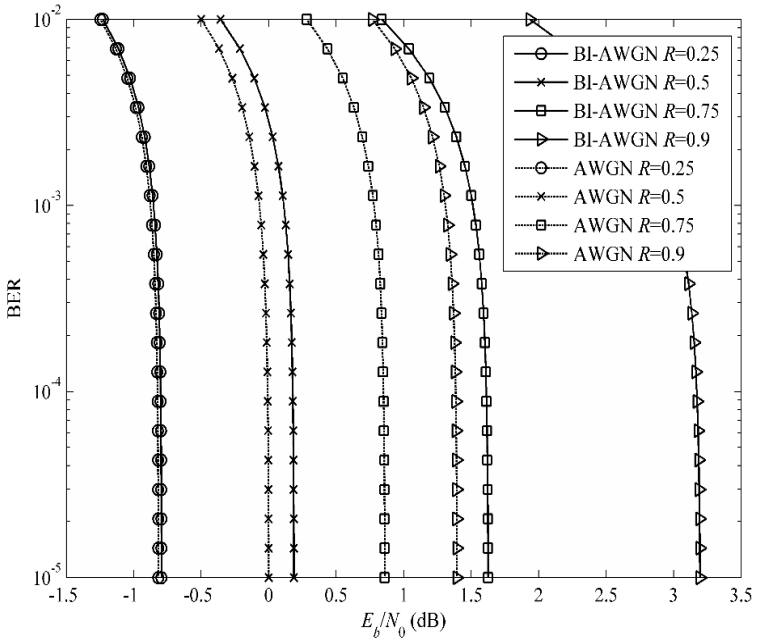


Fig. 2-4 Capacity lower bounds on BER as a function of SNR

2.4 Channel Polarization and Polar Codes

Channel coding technologies are designed to solve the problem regarding how to transmit information reliably through channels from source to destination. Shannon studied this problem from the perspective of mathematical theory, and quantified the maximum amount of information that source can reliably transmit to destination [21], i.e., source transmits information to destination with an arbitrarily low error probability.

If an arbitrary channel code of information length M and block length N in a discrete memoryless channel $(\mathcal{X}, \mathcal{Y}, p(y|x))$ - a (M, N) code consists of the following components:

1) information set $\mathcal{M} = \{1, \ldots, M\}$ with cardinality M;

2) encoder $f : \mathcal{M} \to \mathcal{X}^N$;

3) decoder $g : \mathcal{Y}^N \to \mathcal{M}$.

Then, the coding rate of such a code is defined as

$$R = \frac{\log M}{N}. \tag{2-18}$$

The average decoding error probability is defined as

$$P_e^N = \frac{1}{M} \sum_{i=1}^{M} \Pr\{g(Y^N) \neq i \mid X^N = f(i)\}, \tag{2-19}$$

where X^N denotes the vector $X^N = (X_1, \ldots X_N)$ and Y^N is the vector $Y^N = (Y_1, \ldots Y_N)$. Equation (2-19) indicates the probability of the decoder making error decisions when all possible values in $\mathcal{M}$ are uniformly selected.

If the code rate R is considered achievable, then there would exist a codeword sequence $(\lceil 2^{NR_N} \rceil, N)$, which for any arbitrary $\epsilon > 0$ satisfies

$$\liminf_{N \to \infty} R_N > R - \epsilon, \tag{2-20}$$

$$\lim_{N\to\infty} P_e^N < \epsilon . \tag{2-21}$$

Shannon channel coding theorem proved the existence of very good codes, but it did not provide any specific construction method. The proof of the theorem shows that a lower error probability demands longer random coding, which will increase the probability of getting very good codes, but the complexity of the ML decoding of random codes will also be higher, which could even make decoding impossible. Therefore, one of the main research objectives for channel coding theory has long been to find a practical very good code (a coding method with relatively low encoding and decoding complexity and very good error correction performance), making it feasible to approach the Shannon limit as close as possible with arbitrarily small error probabilities.

2.4.1 Channel Polarization Theory

Consider a B-DMC $(\mathcal{X},\mathcal{Y},W(y|x))$ while $\mathcal{X}$ and $\mathcal{Y}$ represent the input and the output alphabets of the channel, respectively. $W(y|x)$ represents the transition probability of the channel, where $x\in\mathcal{X}$, $y\in\mathcal{Y}$, and the input alphabets $\mathcal{X}$ take value from the set $\{0,1\}$ while the output alphabets and the transition probability have arbitrary values. For this channel, let W^N denote the N uses of the channel W, and then the channel $W^N:\mathcal{X}^N\to\mathcal{Y}^N$ has the following relations:

$$W^N(y_1^N|x_1^N)=\prod_{i=1}^{N}W(y_i|x_i), \tag{2-22}$$

where x_1^N denotes the vector $x_1^N=(x_1,\dots x_N)$ and y_1^N is the vector $y_1^N=(y_1,\dots y_N)$.

For a B-DMC, if there is a permutation relation $\pi:\mathcal{Y}\to\mathcal{Y}$, then all $y\in\mathcal{Y}$ meet the criteria $\pi=\pi^{-1}$ and $W(y|0)=W(\pi(y)|1)$, and the B-DMC is said to be symmetric.

Definition 2-1 (Symmetric Capacity of the B-DMC):

$$I(W)=\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}}\frac{1}{2}W(y|x)\log\frac{W(y|x)}{\frac{1}{2}W(y|0)+\frac{1}{2}W(y|1)}, \tag{2-23}$$

where $I(W)$ denotes the mutual information between input and output symbols of W, in which input symbols are uniformly distributed, and $I(W)\in(0.1)$.

As can be derived from Shannon coding theorem, when the error rate is arbitrarily small, any information transmission rate smaller than the channel capacity is achievable, but how to obtain the closed-loop expression of BER remains unknown. However, for the end-to-end transmission in a B-DMC, it is relatively easy to determine the upper bound on error probabilities when, e.g. the ML decoding method is adopted for decoding. Therefore, we utilizes Bhattacharyya parameters to determine the upper bound on error probabilities [63].

Definition 2-2 (Bhattacharyya Parameters): Bhattacharyya parameters can be expressed as

$$Z(W)=\sum_{y\in\mathcal{Y}}\sqrt{W(y|0)W(y|1)}, \tag{2-24}$$

where $Z(W)$ denotes the upper bound on error probabilities of ML decoding when 0 or 1 is sent through the channel W, where $Z(W)\in(0,1)$.

Judging from the two parameters defined in Definition 2-1 and Definition 2-2, when and only when $Z(W)$ approaches 1, $I(W)$ approaches 0, and vice versa. For an arbitrary B-DMC W, the relational expression between $I(W)$ and $Z(W)$ is strictly as follows [63]:

$$I(W)+Z(W)\geq 1, \tag{2-25}$$

$$I(W)^2+Z(W)^2\leq 1. \tag{2-26}$$

Although these two parameters are simple in form, they play a very important role in the proving of channel coding theory, e.g., proving the existence of channel capacity and the achievability of coding rates [123].

Channel polarization is an outstanding characteristic for constructing coding sequences which can achieve the symmetric capacity of the B-DMC. The concept of channel polarization was later extended to be applied to B-DMCs with arbitrary inputs [66]. The channel codes constructed based on this characteristic is called the polar code.

Suppose the channel $W:\mathcal{X}\to\mathcal{Y}$ is a B-DMC, where $\mathcal{X}=\{0,1\}$. Consider a random vector U_1^2 uniformly distributed in the field $\mathcal{X}=\{0,1\}$, as shown in Fig. 2-5 [63]. Let the input of two independent channels W be

$$X_1^2 = U_1^2 G_2, \tag{2-27}$$

where

$$G_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \tag{2-28}$$

is the kernel matrix for constructing polar codes [70]. Correspondingly, Y_1^2 is the channel output. The channel between U_1^2 and Y_1^2 is defined by the transition probability as follows:

$$W_2(y_1^2 \mid u_1^2) = \prod_{i=1}^{2} W(y_i \mid x_i) = \prod_{i=1}^{2} W(y_i \mid (u_1^2 G_2)_i). \tag{2-29}$$

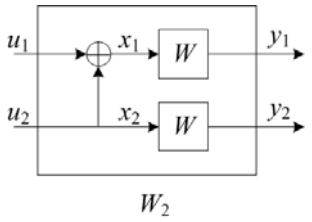


Fig. 2-5 Channel combining of two channels to the synthesized channel W_2

The equation above illustrates the concept of channel combining proposed by Arikan. Its physical significance shows that two independent channels W can be combined, and result in a new synthesized channel $W_2:\mathcal{X}^2\to\mathcal{Y}^2$. It should be noted that there exists a bijection linear conversion between U_1^2 and X_1^2, and all inputs of

the channel W are independent and uniformly distributed and hence we get

$$I(U_1^2;Y_1^2) = I(X_1^2;Y_1^2) = I(X_1;Y_1) + I(X_2;Y_2) = 2I(W). \quad (2\text{-}30)$$

In addition, by using the chain rule of mutual information expressed by Equation (2-5), $I(U_1^2;Y_1^2)$ in the above equation can be separated into

$$\begin{aligned} I(U_1^2;Y_1^2) &= I(U_1;Y_1^2) + I(U_2;Y_1^2 \mid U_1) \\ &= I(U_1;Y_1^2) + I(U_2;Y_1^2,U_1), \end{aligned} \quad (2\text{-}31)$$

where $I(U_1;Y_1^2)$ denotes the mutual information between U_1 and Y_1^2, in which U_2 is regarded as noise, $I(U_2;Y_1^2,U_1)$ is the mutual information between U_2 and Y_1^2, with U_1 in the decoder given.

It should be noted that since U_1 and U_2 are independent, the second equality in Equation (2-31) holds. In addition, the two terms in the equation accurately describe the symmetric capacities of the split channels, i.e. $I(W_2^{(1)}) = I(U_1;Y_1^2)$ and $I(W_2^{(2)}) = I(U_2;Y_1^2,U_1)$. These two types of channels have the following transition probabilities respectively:

$$W_2^{(1)}(y_1^2 \mid u_1) = \sum_{u_2} \frac{1}{2} W_2(y_1^2 \mid u_1^2) = \sum_{u_2} \frac{1}{2} W(y_1 \mid u_1 \oplus u_2) W(y_2 \mid u_2), \quad (2\text{-}32)$$

$$W_2^{(2)}(y_1^2, u_1 \mid u_2) = \frac{1}{2} W_2(y_1^2 \mid u_1^2) = \frac{1}{2} W(y_1 \mid u_1 \oplus u_2) W(y_2 \mid u_2), \quad (2\text{-}33)$$

where $W_2^{(1)}$ denotes the transition probability $W_2^{(1)} : \mathcal{X} \to \mathcal{Y}^2$ and $W_2^{(2)}$ denotes the transition probability $W_2^{(2)} : \mathcal{X} \to \mathcal{Y}^2 \times \mathcal{X}$.

Equation (2-32) and Equation (2-33) illustrate the concept of channel splitting proposed by Arikan. Its physical significance shows that a synthesized channel can be separated into two split channels $W_2^{(1)}$ and $W_2^{(2)}$. The mutual information and

Bhattacharyya parameters of split channels $W_2^{(1)}$ and $W_2^{(2)}$ have the following properties [63]:

$$I(W_2^{(1)}) + I(W_2^{(2)}) = 2I(W), \tag{2-34}$$

$$I(W_2^{(1)}) \le I(W) \le I(W_2^{(2)}), \tag{2-35}$$

$$Z(W_2^{(1)}) \le 2Z(W) - Z(W)^2, \tag{2-36}$$

$$Z(W_2^{(2)}) \le Z(W)^2. \tag{2-37}$$

As can be seen from Equation (2-35), after being split, the channel $W_2^{(1)}$ has poorer performance than the channel W while the channel $W_2^{(2)}$ has better performance than the channel W, which is the most typical example of channel polarization.

Likewise, a similar method can be used to combine two independent channels W_2 so as to generate a new channel $W_4 : \mathcal{X}^4 \to \mathcal{Y}^4$, as shown in Fig. 2-6 [63]. Two independent channels W_4 can also generate $W_8 : \mathcal{X}^8 \to \mathcal{Y}^8$, and so on. In Fig. 2-6, R_4 is the permutation operation which maps the input symbol (s_1, s_2, s_3, s_4) into $v_1^4 = (s_1, s_3, s_2, s_4)$.

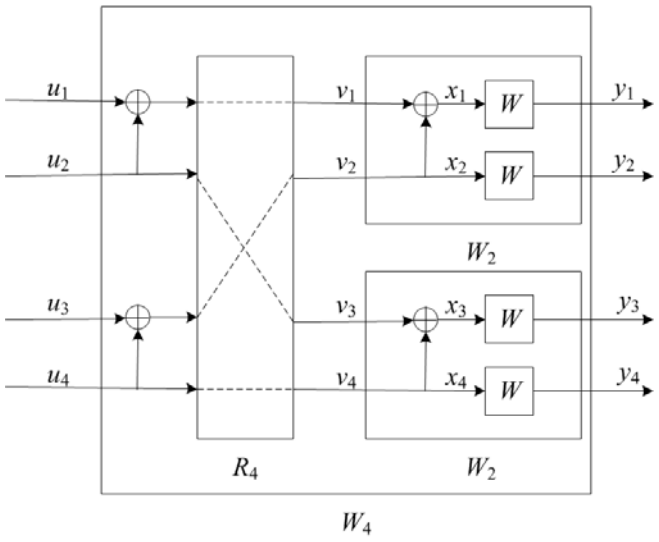


Fig. 2-6 The channel W_4 and its relation to W_2 and W

In general, this kind of channel combining operation can be expressed as a recursive transformation. For an arbitrary $N = 2^n$, define the channel $W_N : \mathcal{X}^N \to \mathcal{Y}^N$ as follows:

$$W_N(y_1^N \mid u_1^N) = W_{N/2}(y_1^{N/2-1} \mid u_{1,e}^N \oplus u_{1,o}^N) W_{N/2}(y_{N/2}^N \mid u_{1,o}^N), \tag{2-38}$$

where $u_{1,e}^N$ denotes the even index of the vector u_1^N and $u_{1,o}^N$ is the odd index of the vector u_1^N. Then, the channel $W_{N/2}$ can also be defined through the channel $W_{N/4}$. Repeat this operation until the expression of the channel W_2 is as shown in Equation (2-29), where $W = W_1$.

Fig. 2-7 [63] illustrates the recursive construction of the channel W_N from two independent channels $W_{N/2}$. Firstly, convert the vector input u_1^N of the channel W_N into s_1^N and then when $1 \le i \le N/2$, relations $s_{2i-1} = u_{2i-1} \oplus u_{2i}$ and $s_{2i} = u_{2i}$ can be obtained. In Fig. 2-7, the operation R_N is a kind of permutation operation called "reverse shuffle". The operation permutes the input s_1^N into a sequence with an output of $v_1^N = (s_1, s_3, \ldots, s_{N-1}, s_2, s_4, \ldots, s_N)$ and then feeds them respectively into two corresponding channels $W_{N/2}$.

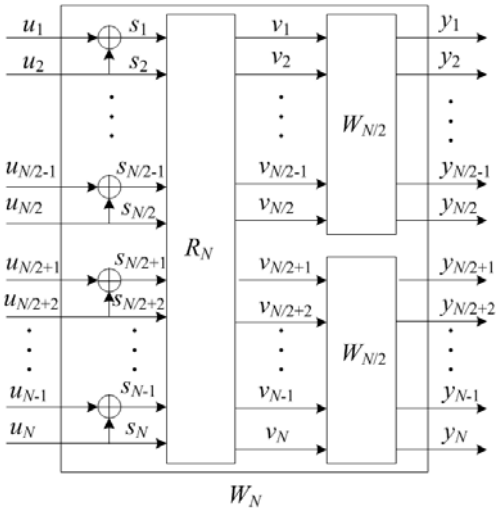


Fig. 2-7 Recursive construction of W_N from two copies of $W_{N/2}$

It can be seen from Fig. 2-7 that $u_1^N \mapsto v_1^N$ is a linear mapping based on a binary

field. Then it can be deduced that the overall mapping relation $u_1^N \mapsto x_1^N$ from the input of the synthesized channel W_N to the input of the channel W^N is also a linear mapping based on a binary field, and can be represented by the generator matrix G_N, and then

$$X_1^N = U_1^N G_N = U_1^N B_N G_2^{\otimes n}, \tag{2-39}$$

where B_N denotes the bit-reversal permutation matrix and $^{\otimes n}$ is the nth order Kronecker product [63].

For all $y_1^N \in \mathcal{Y}^N$ and $u_1^N \in \mathcal{X}^N$, the transition probability between the channel W_N and the channel W^N can be expressed as

$$W_N(y_1^N \mid u_1^N) = W^N(y_1^N \mid u_1^N G_N) = \prod_{i=1}^{N} W(y_i \mid (u_1^N G_N)_i), \tag{2-40}$$

where the second equality shows that the channel W is memoryless, i.e.,

$$W^N(y_1^N \mid x_1^N) = \prod_{i=1}^{N} W(y_i \mid x_i). \tag{2-41}$$

Similar to those in Equation (2-32) and Equation (2-33), the split channels of the channel W_N can also be defined. The splitting operation is computed from the chain rule hereinbefore, as follows:

$$I(U_1^N; Y_1^N) = \sum_{i=1}^{N} I(U_i; Y_1^N, U_1^{i-1}), \tag{2-42}$$

where $I(U_i; Y_1^N, U_1^{i-1})$ denotes the mutual information between U_i and (Y_1^N, U_1^{i-1}).

Define the channel $W_N^{(i)} : \mathcal{X} \to \mathcal{Y}^N \times \mathcal{X}^{i-1}$ as a split channel, and then its transition probability is

$$W_N^{(i)}(y_1^N, u_1^{i-1} \mid u_i) = \sum_{u_{i+1}^N \in \mathcal{X}^{N-i}} \frac{1}{2^{N-1}} W_N(y_1^N \mid u_1^N). \tag{2-43}$$

Based on the construction characteristics of Equation (2-40) for channel combining

and Equation (2-43) for channel splitting, the recursive relation between the channel sets $\{W_N^{(i)}\}$ and $\{W_{N/2}^{(i)}\}$ can be obtained. For an arbitrary $n \geq 0$, $N = 2^n$ and $1 \leq i \leq N$, $W_N^{(i)}$ and $W_{N/2}^{(i)}$ satisfy the relations [63]:

$$W_{2N}^{(2i-1)}(y_1^{2N}, u_1^{2i-2} \mid u_{2i-1}) = \sum_{u_{2i}} \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} \mid u_{2i-1} \oplus u_{2i}) W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2} \mid u_{2i}), \tag{2-44}$$

$$W_{2N}^{(2i)}(y_1^{2N}, u_1^{2i-1} \mid u_{2i}) = \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} \mid u_{2i-1} \oplus u_{2i}) W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2} \mid u_{2i}). \tag{2-45}$$

Meanwhile, $I(W_N^{(i)})$ and $Z(W_N^{(i)})$ have the following relations [63]:

$$I(W_N^{(2i)}) + I(W_N^{(2i+1)}) = 2I(W_{N/2}^{(i)}), \tag{2-46}$$

$$I(W_N^{(2i)}) \leq I(W_{N/2}^{(i)}) \leq I(W_N^{(2i+1)}), \tag{2-47}$$

$$Z(W_N^{(2i)}) \leq 2Z(W_{N/2}^{(i)}) - Z(W_{N/2}^{(i)})^2, \tag{2-48}$$

$$Z(W_N^{(2i+1)}) \leq Z(W_{N/2}^{(i)})^2. \tag{2-49}$$

Based on the above analysis, Arikan continually repeated channel combining and channel splitting in order to polarize channels, and arrived at the channel polarization theorem [63-66].

For an arbitrary B-DMC W, arbitrary $\delta > 0$, and $N = 2^n$, $\{W_N^{(i)}\}$ is polarized through the following methods:

$$\lim_{N \to \infty} \frac{|\{i \in \{1, 2, \cdots, N\} : I(W_N^{(i)}) \in (1-\delta, 1]\}|}{N} = I(W), \tag{2-50}$$

$$\lim_{N \to \infty} \frac{|\{i \in \{1, 2, \cdots, N\} : I(W_N^{(i)}) \in [0, \delta)\}|}{N} = 1 - I(W), \tag{2-51}$$

where $I(W_N^{(i)})$ denotes the mutual information of the bit channel $\{W_N^{(i)}\}$ and $I(W)$ is the symmetric capacity of W.

The theorem shows that, as the block length increases, almost all channels $\{W_N^{(i)}\}$

will be polarized into the corresponding noise-free channels (the mutual information of the channel $\{W_N^{(i)}\}$ close to 1, or the Bhattacharyya parameters close to 0) or noisy channels (the mutual information of channel $\{W_N^{(i)}\}$ close to 0, or the Bhattacharyya parameters close to 1). Therefore, in this thesis those channels with mutual information close to 1 are called "good" bit channels while the remaining channels are called "bad" bit channels.

Fig. 2-8 illustrates the effects of channel polarization as n increases. The horizontal axis represents the index i of the bit channel after polarization and the vertical axis represents the Bhattacharyya parameter $Z(W_N^{(i)})$ of the bit channel. As the diagram shows, values $Z(W_N^{(i)})$ of most bit channels are approaching 0 or 1 while values $Z(W_N^{(i)})$ of a small number of bit channels are between 0 and 1. In fact, the diagram has demonstrated the tendency for about half of the channel sets to center around 0 and the other half to center around 1.

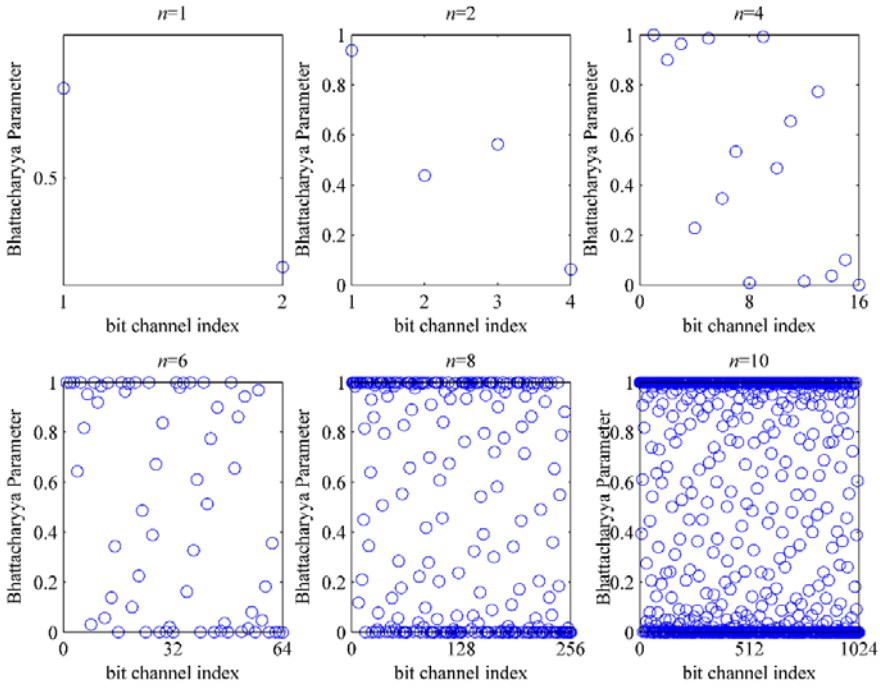


Fig. 2-8 The progress of channel polarization for a BEC with erasure probability 0.5

Of course, there also exist other iterative construction methods which show similar

phenomenon of channel polarization, but even if they can achieve better polarization rates, it is done at the cost of increased encoding complexity [65, 70, 110].

2.4.2 Encoding of Polar Codes

The channel polarization phenomenon illustrates how to construct a polarizable channel and naturally it gives rise to such a channel coding method: information bits are transmitted through the "good" bit channels while frozen and known bit sequences are input into the remaining "bad" bit channels, and are made visible to the decoder in advance. Because the "good" bit channels tends to $I(W)$, it can transmit information reliably at a rate arbitrarily approaching the symmetric capacity. Let $\mathcal{A}$ be a subset of set $[N]=\{1,\ldots,N\}$, and let $u_{\mathcal{A}}$ be a binary vector of length $|\mathcal{A}|$. Then $\mathcal{A}$ can be defined as an information set (i.e., a position index set of the "good" bit channels), and $\mathcal{F}=[N]\setminus\mathcal{A}$ can be defined as a frozen set (i.e., a position index set of the remaining "bad" bit channels apart from that of the "good" bit channels). Accordingly, $u_{\mathcal{A}}$ and $u_{\mathcal{F}}$ can be called information bits and frozen bits respectively. Thus, the encoding method constructed through this channel polarization phenomenon is the polar code.

Definition 2-3 (Polar Codes): For an arbitrary information set $\mathcal{A}\subseteq[N]=\{1,\ldots,N\}$, frozen set $\mathcal{F}=[N]\setminus\mathcal{A}$ and $u_{\mathcal{A}}\in\mathcal{X}^{|\mathcal{A}|}$, let G be the generator matrix G_N defined in Equation (2-39), and let $G_{\mathcal{A}}$ be a submatrix containing partial column vectors of G (the column vector index belongs to the set $\mathcal{A}$), and then the polar code $C_N(\mathcal{A},u_{\mathcal{A}})$ is a linear code, whose codeword set has the following form:

$$C_N(\mathcal{A},u_{\mathcal{A}})=\{x_1^N=u_1^NG_N=u_{\mathcal{A}}G_{\mathcal{A}}\oplus u_{\mathcal{F}}G_{\mathcal{F}}:u_{\mathcal{A}}\in\mathcal{X}^{|\mathcal{A}|}\}. \tag{2-52}$$

Its coding rate is

$$R=\frac{|\mathcal{A}|}{|\mathcal{A}|+|\mathcal{F}|}=\frac{|\mathcal{A}|}{N}. \tag{2-53}$$

As can be seen from the definition of polar codes, the encoding of polar codes firstly fixes the inputs of the bit channels $\{W_N^{(i)} : i \in \mathcal{F}\}$ and then sends information bits through the bit channels $\{W_N^{(i)} : i \in \mathcal{A}\}$. This definition shows that for polar codes constructed by channel polarization, the vector u_1^N contains information bits $u_{\mathcal{A}}$ and frozen bits $u_{\mathcal{F}}$ at the same time. Therefore, to design polar codes for a given channel W to reach a certain error probability, we should determine the coding block length firstly, and then specify the information set and the frozen set respectively. Before a detailed discussion on how to determine the ranges of the sets $\mathcal{A}$ and $\mathcal{F}$, the decoding method for polar codes should be introduced first.

2.4.3 Decoding of Polar Codes

The mutual information $I(U_i; Y_1^N, U_1^{i-1})$ shows that only when both U_1^{i-1} and channel outputs Y_1^N are known can the decoder perform decoding. However, by just knowing those frozen bits, although the decoding of U_i can be carried out, its accuracy cannot be guaranteed. Therefore, Arikan invented the SC decoding algorithm [63]. This algorithm performs decoding in the order of $U_1, U_2, \ldots, U_N$, which ensures that the decoder at least has the estimate of U_1^{i-1} when carrying out the decoding of U_i.

The decision rules of the SC decoding algorithm for polar codes are as follows [63]:

- If $i \in \mathcal{F}$, then set $\hat{u}_i = u_i$;
- If $i \in \mathcal{A}$, then

$$\hat{u}_i = \begin{cases} 0, & L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}) \geq 1 \\ 1, & \text{otherwise} \end{cases}, \tag{2-54}$$

where $L_N^{(i)}(y_1^N, \hat{u}_1^{i-1})$ denotes the likelihood ratio (LR) of u_i when the outputs y_1^N and $\hat{u}_1^{i-1}$ are known, and

$$L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}) = \frac{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1} \mid 0)}{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1} \mid 1)}. \tag{2-55}$$

Based on the ideas of channel combining and channel splitting, the results shown in Equation (2-44) and Equation (2-45) can be used to derive the following recursive iterative formula of the SC decoding algorithm [63]:

$$L_N^{(2i-1)}(y_1^N, \hat{u}_1^{2i-2}) = \frac{L_{N/2}^{(i)}(y_1^{N/2}, \hat{u}_{1,o}^{2i-2} \oplus \hat{u}_{1,e}^{2i-2}) L_{N/2}^{(i)}(y_{N/2+1}^N, \hat{u}_{1,e}^{2i-2}) + 1}{L_{N/2}^{(i)}(y_1^{N/2}, \hat{u}_{1,o}^{2i-2} \oplus \hat{u}_{1,e}^{2i-2}) + L_{N/2}^{(i)}(y_{N/2+1}^N, \hat{u}_{1,e}^{2i-2})}, \tag{2-56}$$

$$L_N^{(2i)}(y_1^N, \hat{u}_1^{2i-1}) = \left[L_{N/2}^{(i)}(y_1^{N/2}, \hat{u}_{1,o}^{2i-2} \oplus \hat{u}_{1,e}^{2i-2}) \right]^{1-2\hat{u}_{2i-1}} L_{N/2}^{(i)}(y_{N/2+1}^N, \hat{u}_{1,e}^{2i-2}). \tag{2-57}$$

When $N/2 = 1$, the recursive formula ends. At this point, there is

$$L_1^{(1)}(y_i) = \frac{W(y_i \mid 0)}{W(y_i \mid 1)}. \tag{2-58}$$

The value of the equation above can be obtained directly from the distribution of the channel W.

It should be noted that the SC decoding algorithm is neither the only nor the best-performing decoding algorithm. In fact, Equation (2-54) shows that if the decoding algorithm can make full use of all the values in the set $\mathcal{F}$ ($i \in \mathcal{A}$, and does not have to satisfy $j < i$) to estimate u_i, a better performance should be achieved. Hussami et al. pointed out that the BP algorithm performs better in decoding polar codes than the SC algorithm does [110]. However, the advantage of the SC decoding algorithm lies in its usage in the theoretical analysis of asymptotic performances. It was based on this algorithm that Arikan proved the achievability of the symmetric capacity of polar codes, and computed the decoding complexity of polar codes to be only $O(N \log N)$.

For an arbitrary B-DMC W and an information set $\mathcal{A}$, when the SC decoding algorithm is employed, the upper bound on the average block error probability of polar codes satisfies [63]:

$$P_e(\mathcal{A}) \le \sum_{i\in\mathcal{A}} \Pr(\hat{u}_i \ne u_i) \le \sum_{i\in\mathcal{A}} Z(W_N^{(i)}) . \tag{2-59}$$

As mentioned earlier, to design polar codes is essentially to determine the rules for the regions of the sets $\mathcal{A}$ and $\mathcal{F}$. Equation (2-59) indicates that the rules can be determined by minimizing the rightmost expression in the equation.

Regarding the average block error probability, Arikan gave a stricter upper bound, i.e., the convergence speed of channel polarization [63, 64]. For a given B-DMC W and an arbitrary $0 < \beta < 1/2$,

$$\lim_{N\to\infty} \Pr(Z(W_N^{(i)}) \le 2^{-N^\beta}) = I(W) . \tag{2-60}$$

The results of the above equation indicate how to choose the frozen set $\mathcal{F}$ when the SC decoder are employed. Based on these results, Arikan proved that polar codes can achieve the symmetric capacity by using the SC decoding algorithm [63].

Theorem 2-1 (Achievability of the Symmetric Capacity of Polar Codes): For a given B-DMC W and a fixed $R < I(W)$, there exists a polar code sequence of block length $N = 2^n$ and rate $R_N > R$. For an arbitrary $0 < \beta < 1/2$ and an arbitrary $n > n_0$, its average block error probability under the SC decoding satisfies

$$P_e(\mathcal{A}) \le \sum_{i\in\mathcal{A}} Z(W_N^{(i)}) = O(2^{-N^\beta}) . \tag{2-61}$$

Proof: For an arbitrary $0 < \beta < 1/2$, let $\beta < \gamma < 1/2$, and choose the information set $\mathcal{A}$ to be

$$\mathcal{A} = \{i \in [N] : Z(W_N^{(i)}) \le 2^{-N^\gamma}\} , \tag{2-62}$$

and then according to Equation (2-60), there is

$$\lim_{N\to\infty} R_N = \lim_{N\to\infty} \frac{|\mathcal{A}|}{N} = I(W) > R . \tag{2-63}$$

For a sufficiently-long block length N, there is

$$N2^{-N^{\gamma}} < 2^{-N^{\beta}}, \tag{2-64}$$

Hence, there is n_0 satisfying

$$P_e(\mathcal{A}) \le \sum_{i\in\mathcal{A}} Z(W_N^{(i)}) < N2^{-N^{\gamma}} < 2^{-N^{\beta}}. \tag{2-65}$$

Because this error probability averages all the frozen sets, there must exist a frozen set, which causes the maximum error probability to be $N2^{-N^{\beta}}$. □

If the channel W is a symmetric channel, the symmetric capacity $I(W)$ equals the channel capacity C. In addition, Arikan pointed out that for a B-DMC, the selection of the value $u_{\mathcal{F}}$ will not influence the average block error probability [63]. Therefore, to simplify the coding design, $u_{\mathcal{F}}$ can be set as 0.

Suppose a polar code of block length $N=16$ is designed for a BEC channel with erasure probability 0.4. Firstly, Bhattacharyya parameters are iteratively computed based on Equation (2-48) and Equation (2-49), as shown in Fig. 2-9. When the polar coding rate is 9/16, 9 bit position indices with minimum Bhattacharyya parameters are selected as the information set $\mathcal{A}$, and then $\mathcal{A}=\{7,8,10,11,12,13,14,15,16\}$ (positions of the bold green figures shown in Fig. 2-9), and correspondingly $\mathcal{F}=\{1,2,3,4,5,6,9\}$ (positions of the bold red figures shown in Fig. 2-9). If the information bit sequence to be sent is $u_{\mathcal{A}}=(1\ 0\ 0\ 1\ 1\ 1\ 0\ 0\ 1)$, let it be passed through the corresponding "good" bit channels in the set $\mathcal{A}$ and set the input of the corresponding "bad" bit channels in the remaining set $\mathcal{F}$ as the frozen bit sequence $u_{\mathcal{F}}=(0\ 0\ 0\ 0\ 0\ 0\ 0)$ so at this point the input coding sequence is (0 0 0 0 0 0 1 0 0 0 1 1 1 0 0 1). Carrying out the channel recursion conversion as shown in Fig. 2-7 on this input coding sequence will generate a codeword sequence $\boldsymbol{x}$ = (1 0 1 0 0 1 0 1 0 0 1 1 0 0 1 1). After having $\boldsymbol{x}$ pass through the BEC channel, the received sequence will be $\boldsymbol{y}$ = (1 ? ? ? 0 ? 0 1 0 0 1 ? 0 ? 1 1)，where ? represents the bits deleted after being

passed through the BEC channel. As the decoder is visible to the corresponding frozen bit sequence in the set $\mathcal{F}$, it is only necessary to apply the SC decoding on the corresponding information bit sequence in the set $\mathcal{A}$. The LR value of information bits can be iteratively computed using Equation (2-56) and Equation (2-57). Lastly, the estimate of the corresponding information bit sequence can be obtained through Equation (2-54), i.e. the SC decoding decision rule.

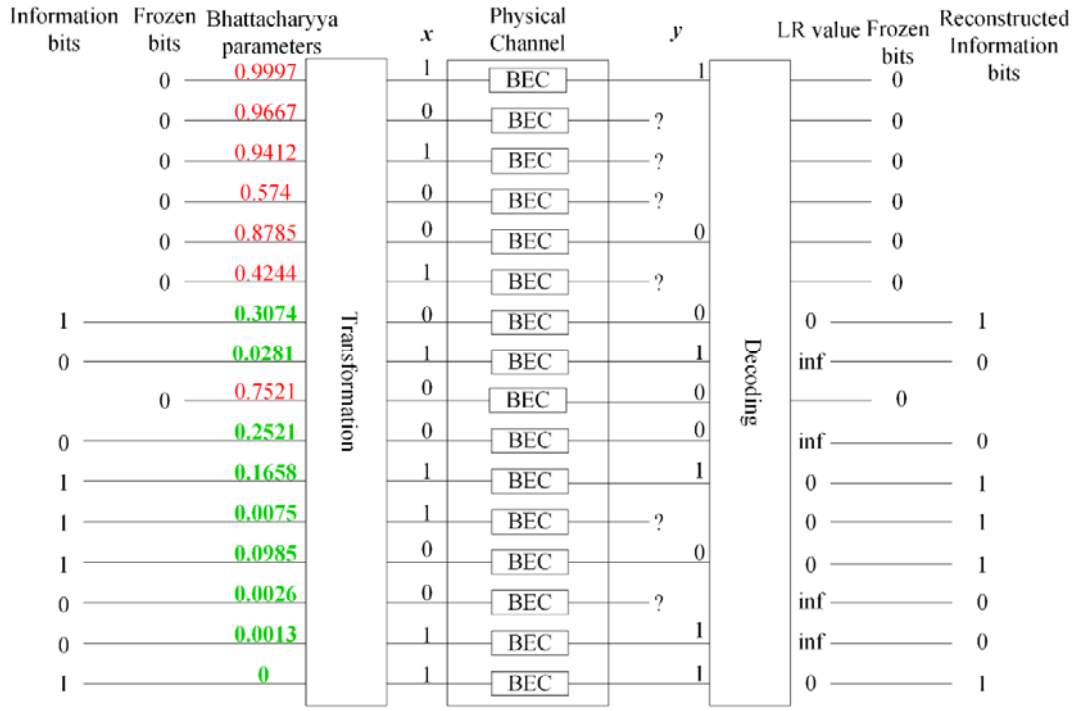


Fig. 2-9 An example of encoding and decoding for a polar code of length 16 with rate 9/16

Although Arikan proved that polar codes can achieve the symmetric capacity, it only serves as evidence of an asymptotic performance. Therefore, it is quite necessary to study the channel polarization rate of polar codes and the SC decoding performance when the block length is finite. Fig. 2-10 [63] is the simulation diagram of the rate versus reliability relation for polar codes of block length $N = \{2^{10}, 2^{15}, 2^{20}\}$ in a BEC channel with erasure probability 0.5. Fig. 2-10 has two curve assemblies, each containing three curves, where the solid lines represent R with respect to $p_U = \sum_{i \in \mathcal{A}} Z(W_N^{(i)})$ and the dashed lines represent R with respect to

$p_L = \max_{i\in\mathcal{A}} Z(W_N^{(i)})$, where R is the rate of polar codes, p_U is the upper bound $P_e(\mathcal{A})$ of block error probability for polar codes with rate R under the SC decoding algorithm and p_L corresponds to the lower bound on $P_e(\mathcal{A})$. By simulating polar codes of finite block lengths, Fig. 2-10 theoretically proves the achievability of the channel capacity of polar codes. As the block length increases, the BER performance curve improves continuously. That means, the faster the curve drops, the closer the performance gets to the theoretical limit. However, it is obvious in Fig. 2-10 that the SC decoding algorithm for polar codes needs to be further improved to obtain a performance comparable to that of LDPC codes.

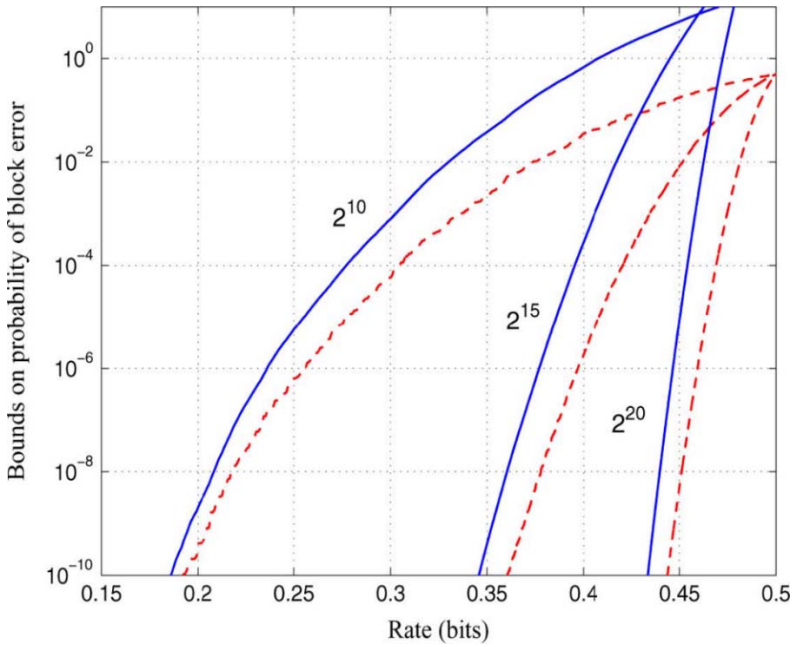


Fig. 2-10 Bounds on block error probability for polar codes with different block lengths

2.4.4 Encoding and Decoding Complexity and Advantages and Disadvantages

One of the main advantages of polar codes is that they enjoy relatively low encoding and decoding complexity, which both are $O(N\log N)$. Let $\chi_e(N)$ be the encoding complexity of the polar code of block length N, and then it can be known from Fig. 2-5 that when $N=2$, only one exclusive-or (XOR) operation is needed to compute $u_0^1 G_2$, so at this point the encoding complexity is $\chi_e(2)=1$. When $N=4$, it can be seen from

Fig. 2-6 that four XOR operations are needed so its encoding complexity is $\chi_e(4)=4$. Similarly, based on the iterative linear transformation relation in Fig. 2-7, there are

$$\chi_e(N)=\frac{N}{2}+2\chi_e(\frac{N}{2}) \tag{2-66}$$

$$=\frac{N}{2}+2(\frac{N}{4}+2\chi_e(\frac{N}{4}))$$

$$=\frac{N}{2}+\frac{N}{2}+4(\frac{N}{8}+2\chi_e(\frac{N}{8}))$$

$$=\ldots=\frac{N}{2}\log N. \tag{2-67}$$

Therefore, Equation (2-67) shows that the encoding complexity of polar codes is only $O(N\log N)$.

Equation (2-56) and Equation (2-57) indicate that the value of $\{L_N^{(i)}\}$ can be computed by using the set $\{L_{N/2}^{(i)}(y_1^{N/2},\hat{u}_{1,o}^{2i-2}\oplus\hat{u}_{1,e}^{2i-2})\}$ and the set $\{L_{N/2}^{(i)}(y_{N/2+1}^{N},\hat{u}_{1,e}^{2i-2})\}$ (where $i\in\{0,\ldots,N/2-1\}$) with its computational complexity being $O(N)$ [63]. Then the computation of the decoding complexity of one polar code of block length N can be simplified into that of two polar codes of block length $N/2$. Suppose the execution of the decoding algorithm is absolutely right and it consumes one unit value, and then the decoding complexity of polar codes of block length $N=2$ is $O(1)$. If we let the decoding complexity of polar codes of block length N be $\chi_d(N)$, then there is

$$\chi_d(N)=O(N)+2\chi_d(\frac{N}{2}). \tag{2-68}$$

Similar to the derivation of Equation (2-67), Equation (2-68) shows that the decoding complexity of polar codes is also $O(N\log N)$.

In summary, the polar code as a brand-new channel coding method has the following advantages:

- Capable of achieving the symmetric capacities of B-DMCs;
- Enjoying both relatively low encoding and decoding complexity;
- Able to arbitrarily change rates without increasing the encoding complexity because of the encoding structure;
- Easy to realize the encoding and decoding structures through hardware and capable of high-rate parallel decoding;
- Free from error floor influence and hence able to be applied in scenarios with strict requirements for error rates, such as wire communications, deep space communications, and disk storage;
- Presently not restricted by intellectual property rights and patents.

Meanwhile, it also has the following shortcomings:

- Still in the early stage of theoretical research with the necessity to further improve the theoretical concepts;
- In need of further improvements in decoding algorithms in order to increase the actual performance of polar codes with finite block lengths;
- Only capable of fully demonstrating its performance advantages when block lengths are long, which leads to a relatively long decoding delay.

2.5 Summary

Polar codes, as a brand new modern channel coding method, not only are the first type of very good codes to have been proven capable of asymptotically achieving the symmetric capacities of B-DMCs, but also boast relatively low encoding and decoding complexity. This chapter, based on channel polarization theory, studied the encoding and decoding algorithm for polar codes and discussed the performance of polar codes from both the theoretical and practical perspectives. With many advantages in coding

construction that are absent from other modern channel coding methods, polar codes have a very good prospect for practical applications in wireless communication systems.

Chapter 3 Encoding and Decoding Algorithms for SCPL codes

3.1 Introduction

Turbo codes and LDPC codes are two types of channel coding methods that are capable of approaching the Shannon capacity limit. However, due to high complexity either in encoding or decoding, both codes need to be further improved in actual applications. In a wireless communication system, encoding and decoding complexity are the key issues that need to be considered. Thus, to identify a channel coding method with good performance in application as well as low encoding and decoding complexity is of great importance in the field of research on channel coding.

As a special type of LDPC codes, LDGM codes have the advantage of low encoding and decoding complexity and are easier to realize [56], and that is why research on the channel coding technology for LDGM codes with low encoding and decoding complexity is of both theoretical and practical significance. LDGM codes are also a type of linear systematic codes, whose generator matrix can be obtained simply through the combination of a sparse matrix and an identity matrix. Therefore, its coding complexity is much smaller than that of LDPC codes. Meanwhile, due to the sparsity of the parity-check matrix, LDGM codes can also adopt the BP decoding algorithm without changing the decoding complexity. Despite the many advantages of LDGM codes, there are a lot of degree-1 variable nodes in their Tanner graph, which makes it impossible to update information through bi-directional transmission, and results in high error floors of LDGM codes independent of the block length [51]. Therefore, LDGM codes are widely considered a coding method with poor asymptotic performance, whose individual application in actual wireless communication systems

is limited to some degree. In 2003, Frias and Zhong proposed a construction method for serially-concatenated LDGM (SCLDGM) codes [56], which significantly alleviates the high error floor problem for the decoding of single LDGM codes, and achieves an excellent performance approaching the Shannon capacity limit [56-60].

In recent years, polar codes have become a key issue for research on channel coding. Compared with modern channel coding schemes such as Turbo codes and LDPC codes, polar codes have the absolute advantages of low encoding and decoding complexity, and being immune from the influence of error floors due to their structural characteristics. Although polar codes have been proven to be able to reach the symmetric capacity of B-DMC when their block lengths tend to infinity, the performance in the decoding of polar codes of finite block lengths needs to be further improved in practical applications as compared with the cases of Turbo codes and LDPC codes. Therefore, Arikan's research results have been extended by a few pieces of literature, which enhanced the decoding performance through the improvement of polar decoding algorithms [71, 72], [112-121]. However, these studies are only concerned with the algebraic structure design of the decoder of polar codes, and barely discussed how to apply them in reality and how to make them approach the Shannon capacity limit.

Currently one of the research focal points on polar codes is to improve their performance in actual applications. In order to enhance the convergence speed of the BER performance of polar codes, Bakshi et al. [124] were the very first to have mentioned the concept of applying concatenated codes. The concatenation involves using a high-rate RS code as the outer code and a short polar code as the inner code. As long as proper design parameters are chosen, the convergence speed of this code-concatenation method will be significantly improved without affecting its

computational complexity. However, this code-concatenation method only proved its asymptotic performance in theory but did not give its practical performance when block lengths are finite. Eslami and Pishro-Nik [115], bearing in mind the actual application of polar codes, designed a code-concatenation method based on polar codes of finite block lengths, which uses polar codes as the outer codes, and LDPC codes as the inner codes (referred to as Polar-LDPC codes in this paper). Its application in real high-rate optical transport networks (OTNs) not only achieved a performance approaching the Shannon capacity limit but also effectively avoided the error floor problem. However, because this code-concatenation method uses LDPC codes as the inner codes, the encoding complexity of the concatenated codes remains high.

Concatenated codes have been widely studied since their invention by Forney in 1966. Up till now, it has been possible to apply concatenated codes in many real and specific communication environments, including deep space communication, long distance optical fiber channel and data storage, etc. To apply channel coding methods in such communication environments is very demanding on the reliability of communication transmission, which requires the BER to be at least as low as 10^{-10}. LDPC codes as an excellent channel coding method can be applied in such communications environments. As SNR increases, its BER performance converges extremely fast within the low SNR region but within the mid-high SNR region, there is a possibility of the abrupt occurrence of error floors before BER falls to 10^{-10}.

From the perspective of practical applications, this chapter firstly introduces the principles of concatenated codes and then it theoretically analyzes the feasibility of serially concatenating polar codes and LDGM codes. According to the analysis, the existence of high error floors for LDGM codes is attributed to large numbers of degree-1 variable nodes found in the Tanner graph. As a result, information being transmitted

between these variable nodes and parity-check nodes can barely get updated. The reason for polar codes being immune from the influence of error floors is analyzed to be that in the Tanner graph of any polar code of a block length equal to or bigger than 8, its girth is at least 12, and the minimum size of the stopping set of polar codes will change with $O(\sqrt{N})$, where N is the block length. Then, in order to effectively address the slow convergence of the BER performance of polar codes and the high error floor problem of LDGM codes, a SCPL coding scheme is proposed which uses polar codes as the outer codes and LDGM codes as the inner codes, thus combining the advantages of both codes. Lastly, this novel channel coding method is tested through simulation with reasonably selected design parameters for the SCPL coding method, such as proper allocation of the coding rates between inner codes and outer codes, and the weights of regular LDGM codes. The simulation results show that the design of serially concatenating the two codes with poor performance in actual application in a proper way makes a very good channel coding method, whose performance extremely approaches the Shannon capacity limit with encoding complexity lower than that of the Polar-LDPC code. The error floor problem is also effectively avoided.

3.2 Principles of Concatenated Codes

In 1966, Forney proposed the concept of concatenated codes as an effective channel coding technology, which constructs short codes into powerful long codes [46]. Concatenated codes are usually composed of two subcodes, which are taken from different numerical ranges and are then serially or parallel concatenated. Concatenated coding schemes, which use non-binary codes (usually RS codes) as the outer codes, and binary codes (usually convolutional codes or Turbo codes) as the inner codes, have already been widely applied in digital communications and data storage. Such coding

schemes reduce the decoding complexity while maintaining a good decoding performance [125]. The outer codes of concatenated codes are usually quite long, which employ the algebraic decoding method; the inner codes are usually short, which use the soft-decision decoding algorithm.

As the concatenation methods for these two types of subcodes differ, this results in different types of concatenated codes. Concatenated codes can generally be divided into serially concatenated codes and parallel concatenated codes. Serially concatenated codes are generated by serially concatenating the two types of subcodes as the inner and outer codes, as shown in Fig. 3-1; and parallel concatenated codes are generated by parallel concatenating the two types of subcodes in an up-down fashion. Therefore, Turbo codes can be regarded as parallel concatenated convolutional codes [125].

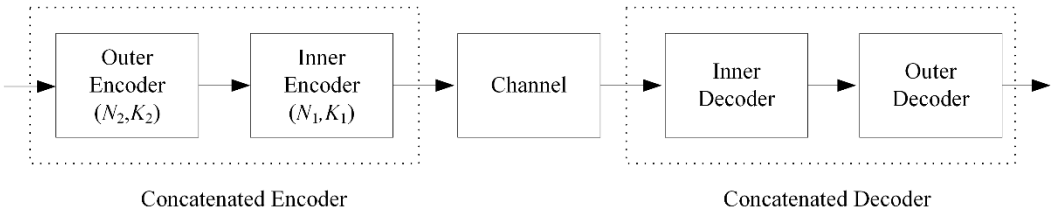


Fig. 3-1 A simple communication system using concatenated codes

The serially concatenated codes shown in Fig. 3-1 consist of two subcodes: A (N_1, K_1) binary code C_I and a (N_2, K_2) non-binary code C_O with its symbol taken from GF(2^{K_1}). The encoding by a concatenated encoder consists of two steps [125]: First, encoding of outer codes, which divides K_1K_2 bits of binary information into K_2 bytes, with each byte containing K_1 information bits. According to the encoding rules of C_O, these K_2 bytes are encoded into a codeword of N_2 bytes. Secondly, encoding of inner codes, which encodes every K_1-bit byte into a codeword in C_I and then generates a codeword constructed by N_2 codewords in C_I of block length N_1N_2. Therefore, the resulting concatenated code is a binary linear block code of

(N_1N_2, K_1K_2). The subcodes C_I and C_O are called the inner code and outer code respectively. To decode the concatenated codes constructed by C_I and C_O, the inner codes and outer codes are decoded in succession, after which the information with error corrected is obtained. Therefore, one important advantage of concatenated codes is that the decoding can be conducted based on each subcode. Because of this advantage, compared with decoders which individually use any subcode, concatenated decoders have significantly lower complexity.

Concatenated codes as a channel coding method can efficiently overcome random errors and burst errors. In case of any small amount of random errors generated after information is passed through a channel, they can be corrected just through the decoding of inner codes C_I alone. However, if random errors or burst errors generated in the channel are large in quantity, they cannot all be corrected by C_I alone because although the errors may have only affected a small number of bytes, it is very likely for them to have already exceeded the correcting capability of C_I. Nonetheless, at this point, a considerable number of errors have already been effectively corrected by C_I. The output errors from the inner decoder are no more than a small input of wrong characters for the outer decoder and can subsequently be easily corrected by C_O [125].

In order to obtain a better decoding performance and at the same time maintain low decoding complexity, short codes are usually used as the inner codes with the soft-decision decoding algorithm while long codes are used as the outer codes with the hard-decision decoding algorithm. If there is no strict requirement for the decoding complexity and the decoding delay, we can consider adopting the soft-decision decoding algorithm for both the inner codes and outer codes and iteratively execute the decoding of the inner codes and outer codes so as to markedly improve the overall

decoding performance. During each iterative decoding process, the soft information output from the inner decoder can be taken as the input of the outer decoder, and then used for the soft-decision decoding of the outer codes. The outer decoder do not make hard-decisions directly on the outputs, but rather exports likelihood ratio soft information and feeds back such soft information to the inner decoder as an input, which will be used for the next iterative decoding. Iterative decoding is continually executed until a certain termination condition is met. As the errors of the inner decoder usually occur in succession, in order to prevent direct correlations among successive iterative decoding, usually an interleaver is required between the inner encoder and the outer encoder. Accordingly, a deinterleaver is required between the inner and outer decoders as well. Concatenated codes employing this kind of iterative decoding have a very good decoding performance, which closely approaches the Shannon capacity limit [125].

So far we have discussed concatenated codes using the serial method but of course concatenated codes can also be realized in a parallel fashion [125]： Through a pseudo random interleaver, two encoders can independently encode the same information sequence, and generate two independent parity-check bits. During the decoding process, two decoders conduct iterative decoding based on these two parity-check bits.

3.3 LDGM Codes and the Analysis of Their Error Floors

3.3.1 LDGM Codes

As a special type of LDPC codes, the encoding structure and the decoding algorithm of LDGM codes are developed on the basis of LDPC codes. That is why LDGM codes possess most of the characteristics of LDPC codes. As shown in Equation (3-1), the parity-check matrix of a systematic LDGM code consists of two parts. The left part is

a randomly structured $K \times (N-K)$-dimensional sparse matrix $\boldsymbol{P}$ while the right one is a K-dimensional diagonal identity matrix $\boldsymbol{I}$. If all the rows (and columns) in $\boldsymbol{P}$ have the same number of "1", this LDGM code is called a regular LDGM code; if all the rows (and columns) in $\boldsymbol{P}$ do not have the same number of "1", then this LDGM code is called an irregular LDGM code.

$$\boldsymbol{H} = [\boldsymbol{P}^T \mid \boldsymbol{I}]$$

$$= \left[\begin{array}{cccccc|cccc} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{array}\right]. \tag{3-1}$$

As the parity-check matrix $\boldsymbol{H}$ has N columns and M rows, a codeword sequence $\boldsymbol{c}$ of N bits must satisfy the parity-check relation $\boldsymbol{H}\boldsymbol{c}^{\mathrm{T}} = 0$. The number of information bits is $K = N - M$, and the coding rate is $R = K / N$.

The parity-check matrix of LDGM codes can also be represented by the Tanner graph [126-128]. The Tanner graph of the parity-check matrix in Equation (3-1) is as shown in Fig. 3-2, in which the variable nodes and check nodes of LDGM codes are respectively denoted by circles and squares. Different from those in the Tanner graph of LDPC codes, the variable nodes of LDGM codes consist of two parts, one corresponding to the systematic information bit nodes, and the other corresponding to the coded bit nodes. Therefore, the only difference between LDGM codes and LDPC codes is that, in the parity-check matrix of LDGM codes, there are $N-K$ degree-1 coded bit nodes with only one edge connected to the parity-check nodes, e.g. $\{s_7, s_8, s_9, s_{10}\}$ as shown in Fig. 3-2. Therefore, LDGM codes cannot update the log-likelihood ratio (LLR) soft information transmitted from the degree-1 variable nodes to the corresponding parity-check nodes.

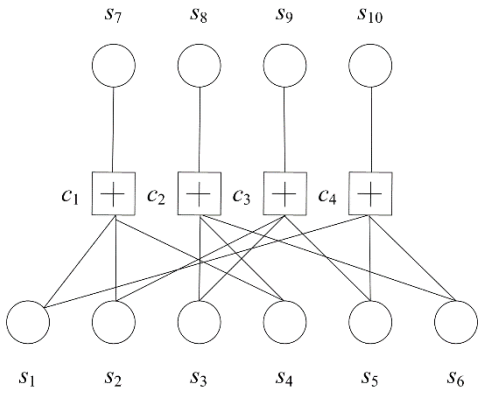


Fig. 3-2 A Tanner graph representation for parity-check matrix of LDGM codes

When systematic information bits, parity-check bits and codeword sequences are respectively expressed as $\boldsymbol{u} = [u_1, u_2, \ldots, u_K]^{\mathrm{T}}$, $\boldsymbol{p} = [p_1, p_2 \ldots, p_M]^{\mathrm{T}}$ and $\boldsymbol{c} = [c_1, c_2, \ldots, c_N] = [u_1, u_2, \ldots, u_K, p_1, p_2 \ldots, p_M]^{\mathrm{T}}$, the encoding of LDGM codes can be represented by the following equation:

$$p_m = \sum_{k=1}^{K} u_k h_{m,k}, \qquad 1 \le m \le N - K \tag{3-2}$$

where $h_{m,k}$ denotes the position corresponding to the *m*th row and *k*th column element in the parity-check matrix.

Because LDGM codes are a systematic code, as can be seen from Equation (3-2), the coded bits of LDGM codewords can be absolutely determined by the sparse parity-check matrix $\boldsymbol{H}$ and can be decoded without converting the parity-check matrix into the corresponding generator matrix, which reduces the encoding complexity to a large degree.

From the above analysis, we know that the parity-check matrices of LDGM codes are just as obviously sparse as those of LDPC codes. Due to such sparsity of the parity-check matrix, the decoding of LDGM codes can also employ the BP decoding algorithm suitable for LDPC codes [129, 130], which only needs to be changed appropriately without increasing the decoding complexity.

3.3.2 Error Floor Analysis of LDGM Codes

To effectively overcome the high error floors of LDGM codes, the cause of such problems must first be analyzed. Similar to how LDPC codes are defined, define (N, K, d_v, d_c) to represent systematic LDGM codes, which have $N-K$ parity-check nodes of degree d_c+1 and K systematic variable nodes of degree d_v, and in which $N-K$ coded bit nodes of degree 1 are connected to the corresponding parity-check nodes, as shown in Fig. 3-2. These degree-1 coded bit nodes are the exact reason for the occurrence of high error floors in systematic LDGM codes. This is because the special structure of LDGM codes causes reliable information to remain unchanged as the initial a priori information input during the transmission process from degree-1 coded bit nodes to parity-check nodes, and not to be influenced by the process of iterative decoding. Therefore, any error occurring in degree-1 coded bit nodes will result in errors in the information being transmitted from the parity-check bit nodes connected to such coded bit nodes to all the systematic variable nodes connected to such parity-check bit nodes. Hence, when errors occur in more than half of the parity-check nodes connected to the systematic variable nodes, errors will occur to the decoding of systematic variable nodes. As the parity-check nodes and the coded bit nodes are in one-to-one correspondence, errors will also occur in more than half of the coded bit nodes and consequently lead to high error floors, which cannot be improved by increasing the block length [51, 56].

To explain it in details, consider a systematic variable node connected to d_v parity-check nodes and a BSC channel with an error transition probability of p. Suppose all the information transmitted to these d_v parity-check nodes from other systematic variable nodes is correct, and d_v is odd, and then if errors occur in at least $(d_v+1)/2$ coded bit nodes of degree 1 connected to these d_v parity-check nodes, the decoding

will fail, causing an erroneous estimate of the information from systematic variable nodes. In the case of d_v being even, if errors occur in more than $d_v/2$ degree-1 coded bit nodes, or errors occur in a total of $d_v/2$ degree-1 coded bits nodes plus systematic variable nodes, then the decoding will produce a wrong estimate for the system variable nodes. Accordingly, the BER of the error floor region can be approximately expressed as [56]:

- When d_v is odd,

$$BER \approx \sum_{k=(d_v+1)/2}^{d_v} \binom{d_v}{k} p^k (1-p)^{d_v-k} . \tag{3-3}$$

- When d_v is even,

$$BER \approx p\binom{d_v}{\frac{d_v}{2}} p^{d_v/2}(1-p)^{d_v/2} + \sum_{k=(d_v/2)+1}^{d_v} \binom{d_v}{k} p^k (1-p)^{d_v-k} . \tag{3-4}$$

As can be known from the above discussion, the drawback of LDGM codes is the existence of high error floors. Although it is possible to design a scheme which makes a compromise between the convergence performance and error floor performance of LDGM codes, a channel coding scheme which solely uses LDGM codes still has its limitations in wireless communication applications.

3.4 BP Decoding and Error Floor Analysis of Polar Codes

3.4.1 BP Decoding Algorithm of Polar Codes

Because of its ability to significantly improve the BER performance of the SC decoding algorithm in practical applications, the BP decoding algorithm is used for the decoding of polar codes in this chapter.

Let the kernel matrix be $G_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, and then $G_2^{\otimes n}$ represents the nth order

Kronecker product of G_2. For example,

$$G_2^{\otimes 3} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{3-5}$$

For an arbitrary $N = 2^n$, $n \geq 1$ and $1 \leq K \leq N$, polar codes are (N, K) linear block codes and their generator matrix is the $K \times N$ dimensional sub-matrix of $G_2^{\otimes n}$, which are generated following the selection rules of information set discussed in Chapter 2.

For a fixed N and all $1 \leq K \leq N$, a diagram describing the transformation relation of the matrix $G_2^{\otimes n}$ can be used to represent polar codes. As shown in Fig 3-3 [131], a transformation diagram of $G_2^{\otimes n}$ when $N = 8$ is described. In the diagram, each edge connected to two nodes carries information of "1" or "0", which is transmitted from the left node to the right node. In each node, modulo-2 addition is performed on all the values from the left node and then the results obtained are delivered through this node to the right node. This diagram describes the linear transformation relation of $x_1^8 = u_1^8 G_2^{\otimes 3}$, where $u_1^8 = (u_1, \ldots u_8)$ and $x_1^8 = (x_1, \ldots x_8)$.

It can also be seen in Fig. 3-3 that the diagram of polar codes has some sparsity, which suggests that the BP decoding algorithm is an effective decoding method for polar codes as well. Based on the idea of Forney-style factor graphs [132], Fig. 3-4 depicts the factor graph of polar codes when $N = 8$ [131].

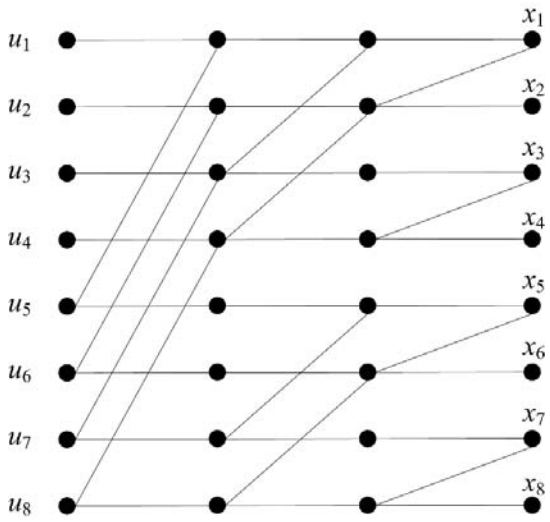


Fig. 3-3 A graph representation for the transformation of $G_2^{\otimes n}$

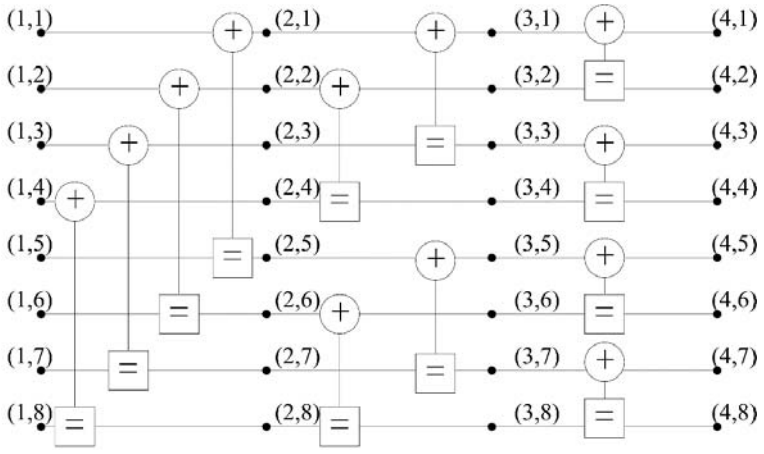


Fig. 3-4 A factor graph representation for the transformation of $G_2^{\otimes n}$

The nodes in the factor graph are marked using the integer pair (i, j), where $1 \le i \le n+1$ and $1 \le j \le N$. From the point of view of the decoder, the leftmost node $(1, j)$ is associated with the source message to be estimated u_j, while the rightmost node $(n+1, j)$ is associated with the channel input variables x_j (i.e., the observed values of the noisy channel output). Apart from these nodes corresponding to these sources and channels, the decoder also needs to estimate whether the value of each node (i, j) is 0 or 1 in the factor graph. The role of the BP decoder is to associate iterative messages through each node (i, j), with the message $R_{i,j}^{(t)}$ being transmitted

to the right and the message $L_{i,j}^{(t)}$ being transmitted to the left ($t = 0,1,\ldots$ is the time index). These messages correspond to the LR values at the time index *t*, so they can be initialized as [131]

$$L_{n+1,j}^{(0)} = \frac{\Pr(x_j = 0 \mid y_j)}{\Pr(x_j = 1 \mid y_j)}, \tag{3-6}$$

$$R_{1,j}^{(0)} = \frac{\Pr(u_j = 0)}{\Pr(u_j = 1)}$$

$$= \begin{cases} 1 & j \in \mathcal{A} \\ \infty & j \in \mathcal{F} \end{cases}. \tag{3-7}$$

All the other $R_{i,j}^{(0)}$ and $L_{i,j}^{(0)}$ are set to 1. It should be noted that when *j* is the corresponding frozen bit index, it is meaningful to set $R_{1,j}^{(0)}$ to ∞ because the decoder has already known that the value of the bit index at the corresponding position in the encoder is 0. On the other hand, when *j* is the corresponding information bit index, to set the value of $R_{1,j}^{(0)}$ to 1 means the a priori information of the 0 or 1 at the bit index position has the same probability.

The basic computational block of BP decoding is a four-terminal processing module, as shown in Fig. 3-5 [131]. Therefore, there are 12 such computational modules in Fig. 3-4. Generally speaking, in the factor graph of polar codes of block length *N*, there are in total $\frac{1}{2} N \log N$ such modules. This computational block module performs the following iterative operations [133]:

$$L_{i,j}^{(t+1)} = f(L_{i+1,j}^{(t)}, L_{i+1,j+N_i}^{(t)} R_{i,j+N_i}^{(t)}), \tag{3-8}$$

$$L_{i,j+N_i}^{(t+1)} = L_{i+1,j+N_i}^{(t)} f(L_{i+1,j}^{(t)}, R_{i,j}^{(t)}), \tag{3-9}$$

$$R_{i+1,j}^{(t+1)} = f(R_{i,j}^{(t)}, L_{i+1,j+N_i}^{(t)} R_{i+1,j+N_i}^{(t)}), \tag{3-10}$$

$$R_{i+1,j+N_i}^{(t+1)} = R_{i,j+N_i}^{(t)} f(R_{i,j}^{(t)}, L_{i+1,j}^{(t)}), \tag{3-11}$$

where $N_i = 2^{n-i}$ and $f(x, y) = (1 + xy) / (x + y)$ for any two reals x and y.

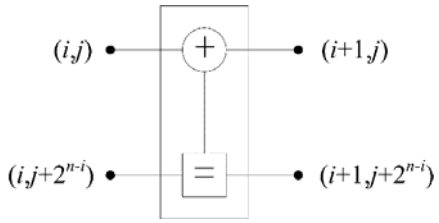


Fig. 3-5 The basic computational block of the BP decoder

3.4.2 Error Floor Analysis of Polar Codes

To analyze the error floors of polar codes, we need to first discuss the girth of the Tanner graph of polar codes. A circle in a Tanner graph refers to a path whose starting and ending points are the same. The length of the smallest circle in the Tanner graph is called the girth of the Tanner graph. When polar codes are decoded using the BP algorithm, the extrinsic information of each node will always remain unrelated to that of other nodes before the number of iterations reaches the girth. Researchers have been committed to constructing codes with a large girth because such codes can achieve a better performance under the BP decoding algorithm.

Theorem 3-1 (Girth of Polar Codes): Any polar code of a block length more than or equal to 8 should have a girth of at least 12.

Proof: As can be seen in Fig. 3-6 [114], there are two types of circles in the Tanner graph of polar codes of $N = 8$: one type as demonstrated by the thick solid line and the other as demonstrated by the thick dashed line. For the first type of circle, its upper and bottom parts are of the same shape in the Tanner graph. Based on the structural characteristics of polar codes, we know that the Tanner graph of polar codes of block length 2^n is contained in that of polar codes of block length 2^{n+1}, which means all the circles in the Tanner graph of polar codes of short block lengths also exist in that of polar codes of long block lengths. See Fig. 3-6 for a polar code of a block length of only 4, whose Tanner graph contains the shortest circle of length 12, including 6

variable nodes and 6 parity-check nodes. On the other hand, the second type of circle in this diagram adopts a structure almost as long as the block length, and thus its range in the diagram is wider. For polar codes of block length 8, even the minimum length of this type of circle is still 12. Based on the above analysis, the girth of the polar code is 12. □

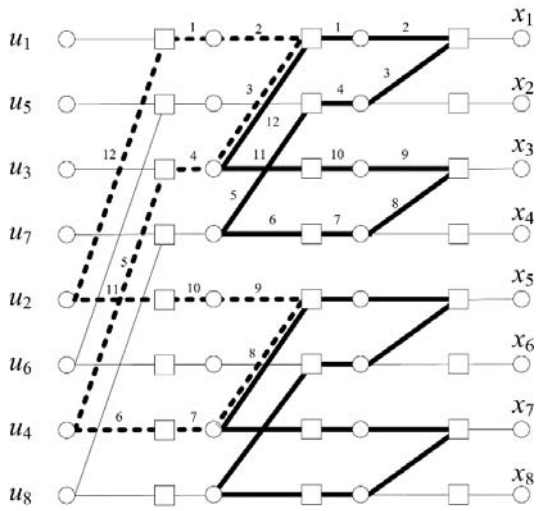


Fig. 3-6 Different types of cycles in the Tanner graph for polar codes with $N = 8$

As we use the BP decoding algorithm for the decoding of polar codes in this section, it is very necessary to study stopping sets (i.e., the stopping distances) and minimal stopping set of polar codes. This is because the stopping sets and the stopping distances play a decisive role in the BER and the error floors of the BP decoding for polar codes. Ref. [114] pointed out that the stopping distances of polar codes vary with the size of $O(\sqrt{N})$ and it is due to such a big influence of the stopping distances that polar codes are less prone to error floors. Besides, according to the simulation results of polar codes in BEC and BI-AWGN channels conducted in Ref. [114], no error floor occurs when BER falls all the way down to 10^{-11}. To sum up, the large girth and stopping distances guarantee that polar codes do not show any error floor.

3.5 Encoding and Decoding Schemes for SCPL Codes

Based on the above analysis, in our proposed scheme, high rate polar codes are used as the outer codes and are serially concatenated with LDGM codes, which in theory should generate a kind of concatenated codes with very good BER performance under different channel conditions. This is because errors occur almost in every frame of the decoding output of LDGM codes but such errors are very small in quantity in each frame. This makes it possible to utilize a type of high rate outer codes, which uses the output of the inner decoder as the a priori probability information to perform the initialization operation. Then through the high rate outer decoder, most frames containing bit errors output from the inner decoder can be further eliminated, which ensures the convergence speed of decoding. Meanwhile, in order to minimize the loss in the outer coding rate, we make the inner coding rate approach the channel capacity as close as possible, and thus the outer coding rate will be close to 1. Lastly, we use polar codes as the outer codes, the output of the outer decoder will not have any error floor. It should be noted that the inner codes and outer codes considered in this paper are all binary codes, which is slightly different from traditional concatenated coding schemes using non-binary codes as the outer codes.

Fig. 3-7 presents the block diagram of SCPL codes proposed in this chapter, which shows that this system serially concatenates binary polar codes, treated as the outer codes, and binary LDGM codes, treated as the inner codes. To be more specific, firstly a polar encoder of rate K / N_1 encodes the information $\boldsymbol{u} = [u_1, u_2, \ldots, u_K]$; then the codeword $\boldsymbol{v} = [v_1, v_2, \ldots, v_{N_1}]$ output from the outer encoder is re-encoded through the LDGM encoder of rate N_1 / N; lastly, the codeword $\boldsymbol{c} = [v_1, v_2, \ldots, v_{N_1}, p_1, \ldots, p_{N-N_1}]$ of overall rate K / N is generated, where $\boldsymbol{p} = [p_1, \ldots, p_{N-N_1}]$ is the coded bit vector of

LDGM codes. After BPSK modulation, the codeword is passed through an AWGN channel with the variance of Gaussian random variable being $\sigma^2 = N_0 / 2$. At the receiver, firstly the LDGM decoder uses the parity-check matrix to perform BP iterative decoding on the received signal $\boldsymbol{y} = [y_1, y_2, \ldots, y_N]$. After several iterative updates of LLR soft information are made between the variable nodes and parity-check nodes in the inner decoder, the LLR soft information (i.e., the soft information output from the corresponding inner decoder) is output at the variable nodes. Then the LLR soft information output from the inner decoder is taken as the a priori probability information of the outer decoder for input initialization. The final LLR soft information (i.e., the soft information output from the corresponding outer decoder) is output after several iterations are made using the BP decoding algorithm for polar codes. Lastly, the decoding results are obtained through decisions.

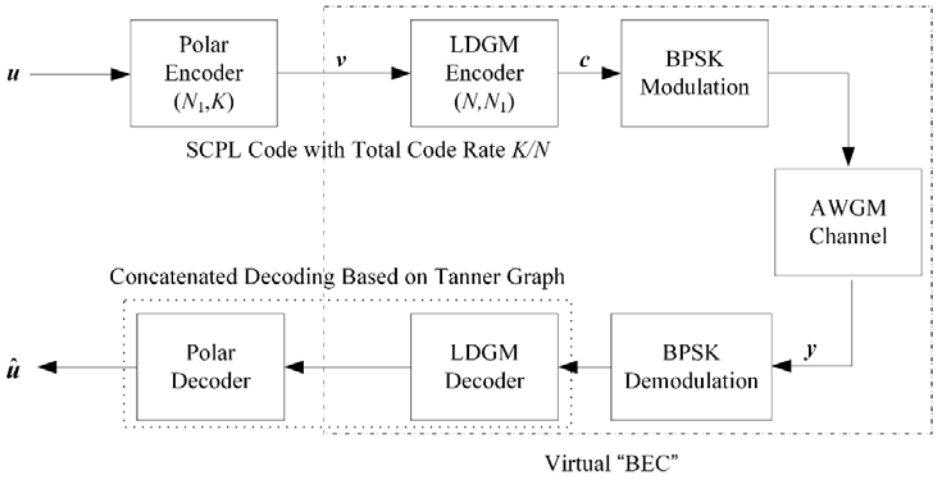


Fig. 3-7 The encoding and decoding block diagram for the proposed SCPL codes

Generally speaking, one problem for LDPC codes is high encoding complexity, which is attributed to the fact that the encoder needs to store a parity-check matrix and a generator matrix of size $O(N^2)$ and consequently the encoding process needs quadratic time. However, in the process of SCPL encoding proposed in this chapter, the encoder only needs to store a parity-check matrix of size $O(N)$ for the inner

LDGM codes, and a generator matrix of size $O(N_1 \log N_1)$ for the outer polar codes, which solves the high complexity problem of Polar-LDPC codes [115] (see Table 3-1).

Table 3-1 Comparison of the complexity between Polar-LDPC codes and SCPL codes

Complexity	Polar-LDPC codes	SCPL codes
Storage size	Parity-check matrix: $O(N)$ Generator matrix N: $O(N_1 \log N_1)$, $O(N^2)$	Parity-check matrix: $O(N)$ Generator matrix： $O(N_1 \log N_1)$
Encoding time	$O(N_1 \log N_1)+O(N^2)$	$O(N_1 \log N_1)+O(N)$

For the decoding of SCPL codes, the BP algorithm is performed in the inner and outer codes sequentially. Their Tanner graph is as shown in Fig. 3-8, where (8,7) polar codes and (13,8) LDGM codes are serially concatenated. $u_2 - u_8$ are the information bits, $v_1 - v_8$ are the coded bit outputs of polar codes (i.e., systematic bits of LDGM codes), $p_1 - p_5$ are the coded bits of LDGM codes and $s_1 - s_5$ corresponds to the parity-check bits of all the degree-1 coded bits. It should be particularly noted that during the encoding process of SCPL codes, the position of the output bit sequence has already been randomly permuted. Therefore, unlike the conventional concatenated codes, SCPL codes do not need interleavers and deinterleavers to correct burst errors.

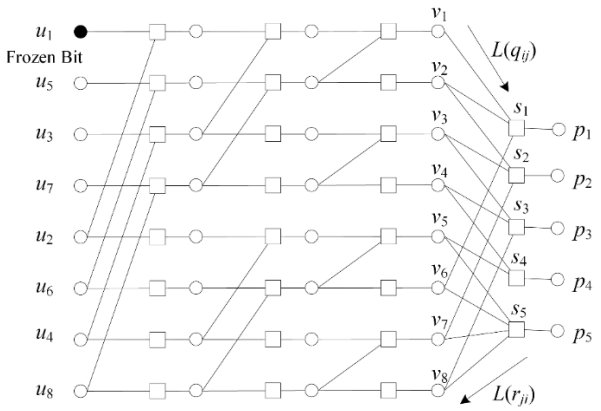


Fig. 3-8 Graph based representation of SCPL codes

From previous discussions, we know that an outer polar decoder can use the output

of the inner LDGM decoder as a priori information to further reduce the decoding error bits from the inner decoder. Hence, the output of the inner decoder already contains the position information of the output error bits. This is because, through simulation, the probability for the inner decoder to correctly output the decoded bit 0 or 1 is close to 1, i.e., the certainty of the inner decoder correctly decoding this bit is 100%. In contrast, the probability for the inner decoder to output the error bit 0 or 1 is close to 0.5, i.e., the certainty of the inner decoder determining the information bit to be 0 or 1 is 50%. This means that, as shown in Fig. 3-7, the section from the output of the outer polar encoder to the output of the inner LDGM decoder can be considered a virtual "BEC" channel and consequently the error information bit outputs from the inner decoder are equivalent to the bits which have already been "erased" in this "BEC" channel. It should be noted that as this "BEC" channel is not a standard BEC channel, one cannot know for sure the specific positions of these "erased" bits. However, since the output of the virtual "BEC" channel contains the a priori probability information of the error bits, it is possible to know where these "erased" bits may exist. The outer decoder treats the output of the inner decoder as the output signals from the BEC channel and then uses the a priori probability information of the output bits from the inner decoder to obtain the a priori information of the positions of the "erased" bits, whose initialized LLR soft information is 0. Finally, the a priori probabilities of all bits are initialized and input into the outer decoder, allowing the outer decoder to further correctly decode the a priori probability information from the inner decoder. Since the channel observed by the outer polar codes is no longer an AWGN channel, but rather a virtual "BEC" channel, the encoding of the outer polar codes should be designed based on the erasure probability ε of the virtual "BEC" channel as follows:

$$\varepsilon = 1 - \frac{K}{N_1}. \tag{3-12}$$

The concatenated decoding algorithm for SCPL codes based on the Tanner graph consists of the following three steps: initializing, decoding of inner LDGM codes and decoding of outer polar codes.

(1) Initializing soft information

At the receiver, firstly the LLR of the received sequence $\boldsymbol{y} = [y_1, y_2, \ldots, y_N]$ is computed as follows:

$$L(c_i) = \log \frac{\Pr(y_i \mid c_i = 0)}{\Pr(y_i \mid c_i = 1)} \tag{3-13}$$

$$= \frac{2y_i}{\sigma^2}, \tag{3-14}$$

where i denotes the ith element in the received signal sequence $\boldsymbol{y}$ and the codeword sequence $\boldsymbol{c}$. Then, the LLR obtained by computation is initialized and separately fed into systematic information bit nodes and parity-check bit nodes as follows:

$$L(q_{ij}) = L(v_i) = L(c_i) \qquad 1 \le i \le N_1, \tag{3-15}$$

$$L(r_{ji}) = L(s_j) = L(c_i) \qquad N_1 + 1 \le i \le N, \tag{3-16}$$

where $L(q_{ij})$ denotes the initialized information of the systematic information bit node v_i and $L(r_{ji})$ denotes the initialized information of the parity-check bit node s_j.

(2) Decoding of inner LDGM codes

i) soft information updates transmitted from s_j to v_i:

$$L(r_{ji}) = 2\tanh^{-1}(L(s_j) \prod_{i' \in V_j, i' \ne i} \tanh(L(q_{i'j} / 2))), \tag{3-17}$$

where $\tanh(x/2) = (e^x - 1)/(e^x + 1)$ and V_j denotes the set that v_i is connected to s_j.

ii) soft information updates passed from v_i to s_j:

$$L(q_{ij}) = L(v_i) + \sum_{j' \in S_i, j' \neq j} L(r_{j'i}), \tag{3-18}$$

where S_i denotes the set that s_j is connected to v_i.

iii) repeat steps i) and ii), until the restricting condition for the stopping of iterations is satisfied, and finally the soft information obtained after the update of the systematic bit nodes is

$$L(Q_i) = L(v_i) + \sum_{j \in S_i} L(r_{ji}). \tag{3-19}$$

If, during the concatenated decoding process, this soft information is used to initialize the bit nodes of the outer polar decoder, then the output errors of the inner LDGM decoder will be further reduced.

(3) Decoding of outer polar codes

Firstly, the soft information $L(Q_i)$ passed from the inner decoder is initialized as

$$L_{n_1+1,i} = L(Q_i), \tag{3-20}$$

where $L_{n_1+1,i}$ denotes the LLR input from the outer polar codes, where $1 \le l \le n_1 + 1$ and $1 \le i \le N_1$.

At the same time, the LLR value of the input information sequence is initialized:

$$R_{1,i} = \begin{cases} 0 & \text{if } l \in \mathcal{A} \\ \infty & \text{if } l \in \mathcal{F} \end{cases}, \tag{3-21}$$

where $\mathcal{A}$ and $\mathcal{F}$ are the information set and the frozen set of polar codes, respectively.

Then the initialized LLR is engaged in computing the updated information in the BP iterative computational block, as follows:

$$L_{l,i} = f(L_{l+1,i}, L_{l+1,i+N_l} + R_{l,i+N_l}),$$

$$L_{l,i+N_l} = L_{l+1,i+N_l} + f(L_{l+1,i}, R_{l,i}),$$

$$R_{l+1,i} = f(R_{l,i}, L_{l+1,i+N_l} + R_{l+1,i+N_l}),$$

$$R_{l+1,i+N_l} = R_{l,i+N_l} + f(R_{l,i}, L_{l+1,i}), \tag{3-22}$$

where $N_l = 2^{n_1-l}$ and $f(x,y) = \ln((1+xy)/(x+y))$ for any reals x and y.

Finally, through the decoding decision rule

$$\hat{u}_i = \begin{cases} 1, & \text{if} \quad L_{1,i} + R_{1,i} < 0 \\ 0, & \text{otherwise} \end{cases}, \tag{3-23}$$

the estimate of the information bit sequence is obtained. When the estimate of the decoding output is identical with the information delivered, or the number of iterations is reached, the iteration stops; if the maximum number of decoding iterations has been reached, but the estimate of the decoding output is still inconsistent with the information delivered, the decoding fails.

3.6 Simulation Design for SCPL Codes

For SCPL codes, the combination of the coding rates of the inner codes and outer codes has a great impact on the decoding performance. Meanwhile the selected weights of inner LDGM codes will influence the decoding performance of the concatenated codes. Therefore, in order to design a SCPL coding scheme for achieving the best BER performance, the design parameters, i.e., the combination of inner and outer coding rates and the inner code weights, must first be properly determined. In this section, we will construct the SCPL coding scheme through computer simulations. Firstly, the combination of inner and outer coding rates is considered. Assume that two subcodes of SCPL codes are the (N_1, K) polar code and the (N, N_1) LDGM code respectively, the concatenated coding rate R is 0.5 and then the product of the rates of inner codes and outer codes is 0.5. Secondly, to simplify the realization of the SCPL coding scheme, only LDGM codes with regular sparse matrices are considered, i.e. the row weight and the column weight of the sparse matrix are equal and are both w. It

should be noted that in all the simulations, ***P***-matrix is generated by following the progressive edge growth (PEG) algorithm [133], and the information of at least 10^9 frames is simulated. In order to decode each frame, both the inner and outer codes use BP decoding till the correct information bit is estimated or the required number of iterations is reached.

As far as we know, the SCPL coding scheme proposed in this chapter cannot use the analyzing method similar to that of the extrinsic information transfer (EXIT) chart [134] so as to theoretically search for the optimal distribution of inner and outer coding rates. Fortunately, through computer simulations, we can in sequence find the proper allocation of inner and outer coding rates and the weights of the inner codes.

3.6.1 Construction Procedure of SCPL Codes

Define the inner and outer coding rates as R_L and R_P respectively, and then $R = R_L \times R_P = 0.5$. Also select the inner code weight $w = 6$ and outer code block length $N_1 = 2^{12}$ as the simulation parameters. Because using high-rate polar codes as the outer codes can remove most error bits output of the inner codes, we present the BER performance curves of three different combinations of the inner and outer coding rates in Fig. 3-9 as a representative sampling of all the combinations of the inner and outer coding rates. The combinations of the inner and outer coding rates (R_P, R_L) respectively are (0.9804,0.51), (0.9615,0.52) and (0.9434,0.53). It should be noted that both R_L and R_P are larger than the rate of 0.5. To minimize the loss in the outer coding rate, R_L is chosen to be close to R, and then R_P is extremely close to 1. As shown in the diagram, when the coding rate R_P is very large, the performance of SCPL codes is not very good, which is demonstrated by the obvious existence of error floors. However, with the decrease of R_P, the error floors disappear, but at the same time the

BER convergence speed also decreases. Therefore, there is this dilemma about making a compromise between the convergence performance and the error floor. Hence, in the following simulation, the combination of coding rates $(R_P, R_L) = (0.9615, 0.52)$ will be regarded as a compromising simulation parameter.

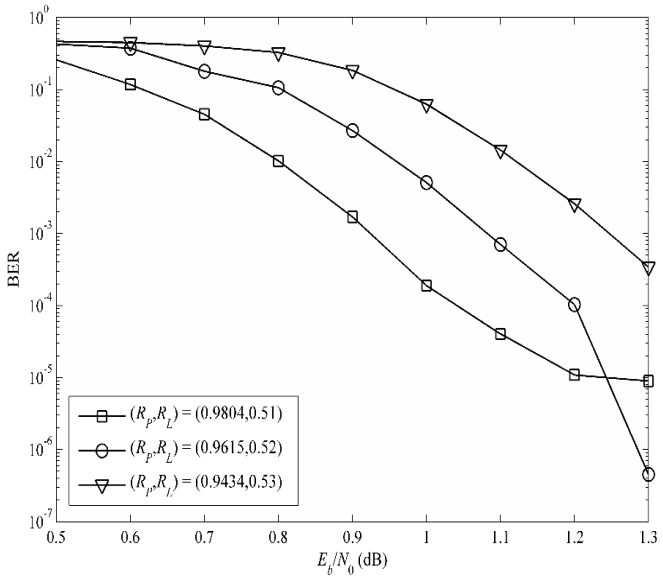


Fig. 3-9 The BER comparison of the SCPL coding scheme with different rate allocations

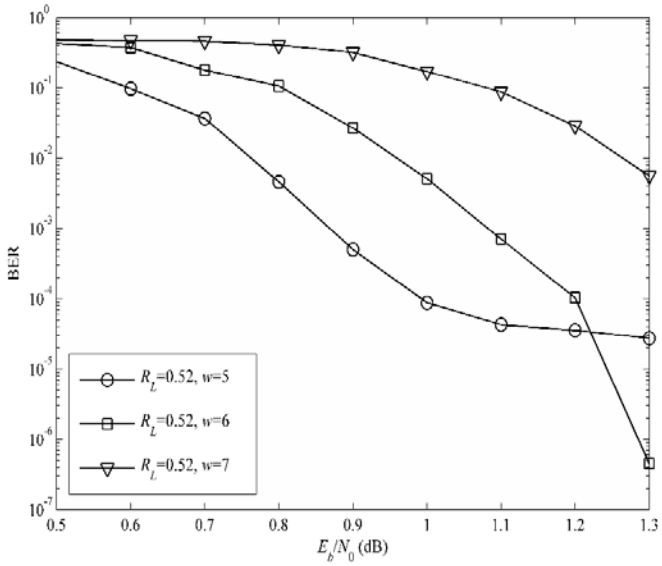


Fig. 3-10 BER comparison of the SCPL coding scheme with different *w*

As the output of the inner decoder has some influence on the a priori probability information input of the outer decoder, we will next analyze how the selected weights of LDGM codes influence the BER performance of the SCPL coding scheme. Fig. 3-

10 shows the BER performance curve of SCPL codes of outer code length $N_1 = 2^{12}$ and inner coding rate 0.52. It can be seen from the diagram that different inner code weights do have a big impact on the performance of SCPL codes. Specifically, even though SCPL codes get the optimal BER performance when $w = 5$, there are obvious error floors. However, with the increase of *w*, the error floors become lower and lower till they disappear eventually. So, in the following simulation, the inner code weight $w = 6$ will be treated as a compromising simulation parameter.

3.6.2 Simulation Results and Performance Comparison

Fig. 3-11 shows the simulation results of the BER performance of SCPL codes when $R_L = 0.52$ and $w = 6$. Not only is the performance curve of SCPL codes depicted, it is also compared with the performances of LDGM codes and polar codes in the AWGN channel, both of coding rate 0.5 and block length 2^{13}. As shown in the diagram, SCPL codes greatly improve the performance of polar codes, and have a fast convergence speed. Meanwhile, they significantly alleviate the error floor problem of LDGM codes. In addition, the diagram also gives the performance curves of LDPC codes and Polar-LDPC codes in [115], where the LDPC codes have a block length of 2^{13} and a pair of irregular degree distribution functions as follows [134]: $\lambda(x) = 0.508x + 0.419x^3 + 0.073x^{17}$ and $\rho(x) = x^7$; the convergence threshold of such LDPC codes is 0.5dB. As the diagram shows, the performance of these SCPL codes is very close to that of Polar-LDPC codes. Also, compared with the case of LDPC codes, error floors do not appear until BER drops to 10^{-10}. Although the performance convergence speed of Polar-LDPC codes is better than that of SCPL codes, they have high encoding complexity.

Fig. 3-12 presents the performance gain curves of SCPL codes of various block lengths. The diagram shows that, as expected, the BER performance gets better and

better as the block length increases. To be more specific, error floors do not appear for the three curves even till BER falls to 10^{-10}. With the channel capacity limit under this scheme being 0.18dB, the diagram shows that when BER=10^{-5} is used as the standard for measuring reliable communication, the performance of SCPL codes of block length $N_1 = 2^{14}$, $R_L = 0.52$ and $w = 6$ is only 0.8 dB away from the Shannon capacity limit. It is believed that with the increase of the block length, the performance of SCPL codes will infinitely approach the theoretical capacity limit.

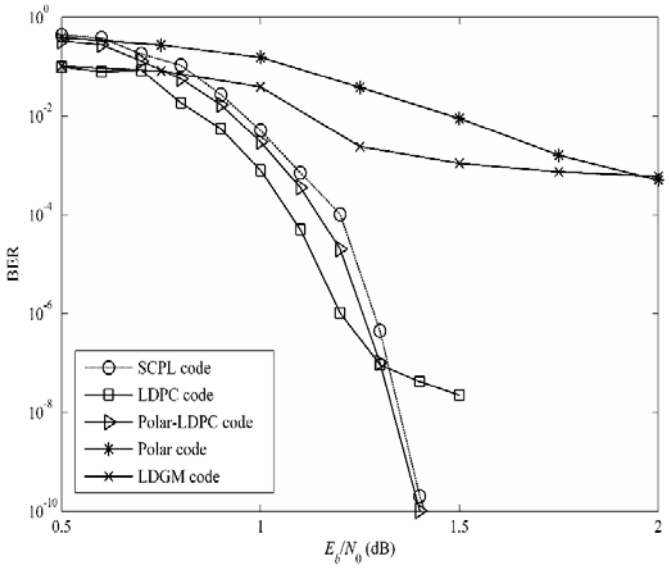


Fig. 3-11 The BER comparison for different coding schemes

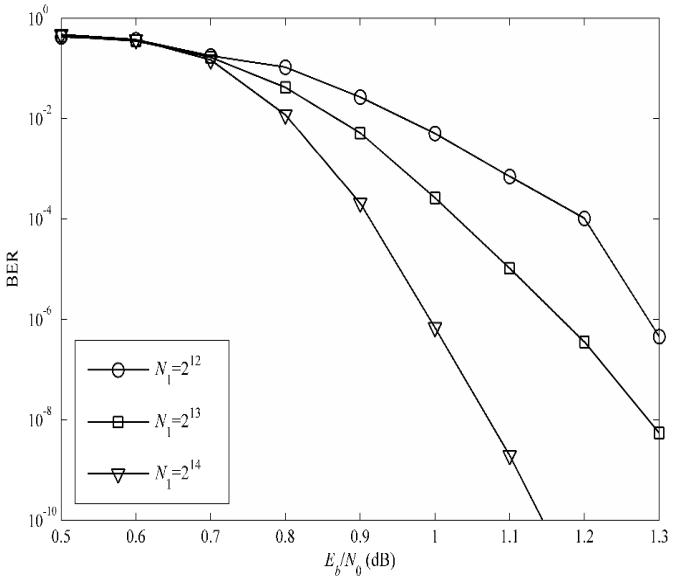


Fig. 3-12 The BER performance of the SCPL coding scheme for different block lengths.

3.7 Summary

In this chapter, we first introduced the basic principles of classic concatenated codes. Then, targeting the slow convergence speed of polar codes, and the high error floors of LDGM codes, we theoretically proved the feasibility of serially concatenating these two channel codes of both low encoding and decoding complexity and then proposed a SCPL coding scheme with low encoding and decoding complexity. Lastly, by properly choosing simulation parameters, we described the method for designing SCPL codes and conducted actual simulation on these novel SCPL codes. The simulation results show that SCPL codes not only have an excellent performance approaching the Shannon theory limit, but also are free from the error floor problem. Therefore, SCPL codes can effectively address the issue regarding the actual application of polar codes and provide a brand-new method for research on channel coding schemes in the fields of deep space communication, long distance optical fiber channel and data storage, etc.

Chapter 4 Polar Coding Schemes for Single-Relay Transmission Systems

4.1 Introduction

With the increasing tension on the resources of wireless spectrum, it has become an issue of common concern to academia and industry both at home and abroad as to how to realize interconnection between more users and support higher-rate and higher-bandwidth wireless data transmission while making full use of the limited spectrum resources. In recent years, cooperative communication technologies, as the key technology for data transmission between wireless network terminals, relay stations (relay terminals) and base stations, have solved the problem regarding high-rate data transmission and large-scale network topologies.

As the most basic component of cooperative communication systems such as cellular mobile communications and wireless Ad Hoc networks, the relay channel model is of great theoretical significance. From the viewpoint of information theory, adding relay nodes between the source node and the destination node to help transmit information from the source node is conducive to increasing communication channel capacity [77]. However, how to use channel coding schemes to improve the reliability of relay systems with the aim of approaching or reaching the channel capacity poses a key problem to current research.

In the relay channel model, relay nodes have two commonly used modes of communication, i.e., half-duplex and full-duplex modes. The half-duplex mode is characterized by relay nodes sending and receiving data only using different frequency bands or during different time phases, while in comparison, in the full duplex mode, relay nodes can send and receive data using the same frequency band or during the

same time phase. As the half duplex mode is easier to be configured in relay equipment at a lower realization cost, this mode is more commonly used in actual wireless device applications. However, the full-duplex mode boasts a higher system capacity [77, 135].

So far, there have been a handful of research findings regarding the application of polar codes in single-relay channels [107-109]. However, all such work was based on the full-duplex mode and/or the hypothesis that the destination node has orthogonal receiver components. Under this hypothesis, by using the properties of orthogonal multiple access, a two-node transmission system can be simplified into a simple end-to-end transmission system, thus avoiding the necessity to study the design of polar coding schemes for scenarios when there is interference between the source signal and the relay signal. However, in actual wireless relay systems, the multiple-access channel model in the system must be considered because in non-orthogonal multiple access mode, it is inevitable that the source signal and the relay signal interference with each other. Therefore, it is necessary to study the polar coding schemes in this mode. In addition, although existing research findings show that there are polar codes capable of achieving the capacity of the relay system, it still remains a challenging topic as to how to design polar codes with low encoding and coding complexity for more practical BI-AWGN channels. By now, there is no relevant research on methods of constructing polar codes in actual half-duplex relay channels.

This chapter firstly presents a system model and then analyzes the upper and lower bounds on the capacity of the half-duplex relay channel. Although distributed channel coding technologies such as convolutional codes, Turbo codes and LDPC codes have been applied to this channel model and have obtained performances approaching the channel capacity limit, there is still not one capable of achieving the capacity limit. Secondly, for the degraded half-duplex single-relay channel model, we propose a

cooperative partial message relaying (CPMR) transmission protocol based on polar codes. Theoretical analysis shows that to apply the CPMR scheme in this channel model can asymptotically achieve the capacity limit of the relay channel and we also derive the upper bound on the average block error probability under the SC decoding algorithm. Then, through the analysis of the channel capacity limit, we elaborate on several key factors affecting the capacity limit. Lastly, according to the multi-access channel model in the relay channel, a JISPIC receiver structure based on polar codes is designed, based on which, we verify through simulation the feasibility of the CPMR scheme of finite block lengths in an AWGN channel employing BPSK modulation. The CPMR scheme proposed in this chapter not only inherits the advantages of the low encoding and decoding complexity of polar codes, but also is able to obtain a performance comparable to that of conventional LDPC coding schemes.

4.2 Half-duplex Single-relay Channel

4.2.1 Channel Model

A discrete memoryless single-relay channel is a simple three-node communication model [13, 77], as shown in Fig. 4-1. The relay channel in the diagram is composed of three nodes, i.e., source node (S), relay node (R) and destination node (D). The link between S and R is called the SR channel, that between S and D the SD channel and that between R and D the RD channel. With the assistance of R, S sends messages to D through the SR channel and the SD channel.

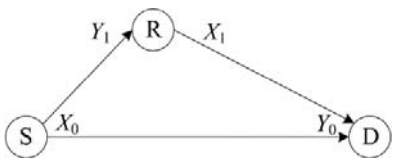


Fig. 4-1 The general model of the single-relay channel

If a relay node transmits information in half duplex mode, it means that it cannot send or receive information in the same frequency band at the same time. In other words, the sending channel and the receiving channel must be orthogonal. This orthogonality can be frequency domain orthogonal, time domain orthogonal, or time-frequency domain orthogonal. If the relay node attempts to send and receive information in the same frequency band at the same time, then received signals will interfere with each other. In theory, as the transmitted signals are known, the relay node can eliminate this interference. However, from the perspective of practical applications, as transmitted signals are usually larger than received signals by 100-150dB [9], this will cause errors to the interference cancellation algorithm. Due to the difficulty in accurately performing the interference cancellation algorithm, wireless communications with full duplex mode have not been widely applied. However, the rapid development of analogue processing technologies gives full duplex mode the potential of being applied in relay channels. Although almost all of the earlier research on relay channel was conducted based on full-duplex relaying, much of the research in recent years, particularly research on actual transmission protocols, was premised on half-duplex relaying.

Take the half-duplex relay channel model as shown in Fig. 4-2 for example. Consider a binary-input discrete memoryless relay channel, and assume that the relay node works in time-division half-duplex mode. In each information block transmission, the overall transmission (normalized) is divided into two phases, i.e., the broadcast (BC) phase and the multi-access (MAC) phase. Let the interval of a BC phase be t, then the interval of a MAC phase is $1-t$. The communication process of this system is as follows: In the BC phase, S broadcasts X_0, and R and D receive Y_1 and Y_0 respectively. At this point, R only receives but does not send. After receiving the

information sent from S, R generates new information X_1 in a certain way (see the brief discussion on cooperative strategies in the next paragraph). In the MAC phase, R sends X_1 but does not receive new information, and meanwhile S sends X_2. Y_2 is the final information received by D in the MAC phase.

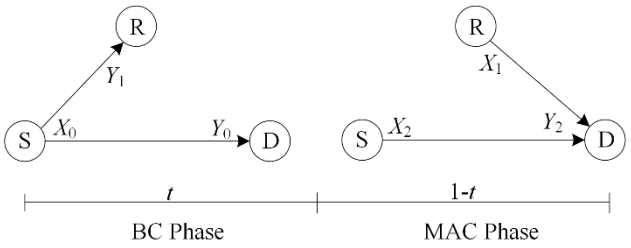


Fig. 4-2 The time-division half-duplex single-relay channel model

There are three most basic and commonly-used cooperative strategies in relay systems [7, 9, 77]. Such cooperative strategies are called relay protocols. The first cooperative strategy is the amplify-and-forward (AF) protocol, under which, the relay node only applies simple signal amplification to received signals and then forward them to the destination node, with the amplification factor being determined by the relative strengths of the SR and SD channels. The major shortcoming of the AF protocol lies in that while amplifying signals, the relay also amplifies noise. The second cooperative strategy is the DF protocol. Under this protocol, the relay node can decode information sent from the source node and after re-encoding the information, forwards the re-encoded information to the destination node. When the SR channel is very good, i.e., when the source node and the relay node are close to each other in physical distance, the DF protocol is close to optimum. When the RD channel is very good, i.e., when the relay node is close to the destination node in distance, usually a third cooperative strategy - the compress-and-forward (CF) protocol will be used. Under this protocol, the relay node does not decode information received from the source node, but rather it independently observes the received information in order to assist the decoding by the destination node. Therefore, it is equivalent to the relay node

sending a piece of estimation information of the source node to the destination node. Such a strategy is also known as estimate-and-forward (EF) or quantize-and-forward (QF).

In this chapter, we mainly consider relay nodes using the DF protocol, i.e., it is assumed that the relay node is relatively close to the source node in distance. The DF protocol satisfying this assumption is superior to other protocols [7] so the transmission between the source node and the relay node is reliable. In order to ensure the reliability of DF cooperative communication systems, researchers have proposed a number of channel coding technologies [14-20]. At present, channel coding technologies for DF cooperative relaying mainly focuses on two aspects:

- Utilizing information theory to analyze the capacities of relay channels;
- Making full use of the relay channel capacity limit and designing efficient distributed channel coding for multiple users in order to further improve the efficiency and reliability of communication systems.

4.2.2 Capacity of Half-duplex Single-relay Channels

Similar to the case of end-to-end communication channels, a basic problem for relay channels is the channel capacity. Many basic definitions of end-to-end communication channels can also be applied to relay channels, such as transmission rate, achievable rate and channel capacity, etc. However, the design of channel coding schemes cannot just be confined to the design of source encoders and destination decoders but it also needs to include the coding design for the relay node. Up till now, the closed-loop expression of relay channel capacity is still unknown and we can only determine its upper bound and lower bounds. Also only under certain conditions will these bounds be tight.

Khojastepour [135], using cut-set theorem for multi-state networks and random

channel coding technologies, derived the upper and lower bounds on the half-duplex relay channel.

The upper bound on the capacity of the half-duplex single-relay channel is

$$C^U \le \sup_{t,p(x_0,x_1,x_2)} \min\{tI(X_0;Y_0,Y_1)+(1-t)I(X_2;Y_2\mid X_1),$$

$$tI(X_0;Y_0)+(1-t)I(X_1,X_2;Y_2)\}. \tag{4-1}$$

The lower bound on the capacity of the half-duplex single-relay channel is

$$C^L \ge \sup_{t,p(x_0,x_1,x_2)} \min\{tI(X_0;Y_1)+(1-t)I(X_2;Y_2\mid X_1),$$

$$tI(X_0;Y_0)+(1-t)I(X_1,X_2;Y_2)\}. \tag{4-2}$$

If the transition probability of the relay channel can be written as follows:

$$p(y_0,y_1,y_2\mid x_0,x_1,x_2)=p(y_1\mid x_0)p(y_0\mid y_1)p(y_2\mid x_1,x_2), \tag{4-3}$$

then the relay channel is called physically degraded and the lower bound on the capacity of the half-duplex relay channel is the capacity of the physically degraded half-duplex relay channel, shown as follows:

$$C=\sup_{t,p(x_0)p(x_1,x_2)} \min\{tI(X_0;Y_1)+(1-t)I(X_2;Y_2\mid X_1),$$

$$tI(X_0;Y_0)+(1-t)I(X_1,X_2;Y_2)\}. \tag{4-4}$$

Khojastepour [135] and Chakrabarti [18] pointed out that when x_0, x_1 and x_2 are uniformly distributed and the time allocation factor t ($0\le t\le 1$) has taken the optimum, the relay channel capacity in Equation (4-4) is achievable. At this point, the two terms in the expression on the right-hand side of Equation (4-4) are equal.

If the half-duplex relay channel is only stochastically degraded, there exists a distribution $q(y_0\mid y_1)$, which makes $p(y_0,y_2\mid x_0,x_1,x_2)$ satisfy

$$p(y_0,y_2\mid x_0,x_1,x_2)=\sum_{y_1} p(y_1\mid x_0)q(y_0\mid y_1)p(y_2\mid x_1,x_2). \tag{4-5}$$

At this point, the expression in Equation (4-4) only represents the achievable DF rate.

It should be noted that, as can be seen from Equation (4-3) and Equation (4-5), physical degradation implies stochastic degradation.

4.3 CPMR Transmission Strategy of Infinite Block Lengths

Based on the polarization phenomenon of polar codes, this section presents a low complexity CPMR protocol, with the aim of achieving the capacity of degraded half-duplex relay channels.

4.3.1 CPMR Transmission Strategy in BC Phase

If the SD channel in Fig. 4-2 is expressed as W_{SD}, the SR channel as W_{SR}, and it satisfies that W_{SD} is the stochastically degraded channel (expressed as $W_{SD} \preceq W_{SR}$) of W_{SR}, then we get the following lemma.

Lemma 4-1: let W_{SD} and W_{SR} be two B-DMCs and satisfy $W_{SD} \preceq W_{SR}$. Meanwhile let $W_{SD}^{(i)}$ and $W_{SR}^{(i)}$ be the ith polarized bit channel generated by polar codes based on the channels W_{SD} and W_{SR}, in which $i \in [N]$. Following Equation (2-62), define the information sets of W_{SD} and W_{SR} as $\mathcal{A}_{SD}$ and $\mathcal{A}_{SR}$ respectively, then we have the following characteristics:

1) the Bhattacharyya parameters of the channels $W_{SD}^{(i)}$ and $W_{SR}^{(i)}$ satisfy [65]:

$$Z(W_{SD}^{(i)}) \geq Z(W_{SR}^{(i)}); \tag{4-6}$$

2) if $i \in \mathcal{A}_{SD}$, then $i \in \mathcal{A}_{SR}$, which means [109]

$$\mathcal{A}_{SD} \subseteq \mathcal{A}_{SR}. \tag{4-7}$$

From the polar code construction methods for the two channels in Lemma 4-1, we know that all the "good" bit channels in W_{SD} are also "good" bit channels in W_{SR} (i.e., Equation (4-7)). Therefore, we define all the bit channel index sets of the source node

in the BC phase as $[N_{BC}]$, then $[N_{BC}]$ can be divided into three subsets corresponding to the bit channels and yet independent and disjoint from each other, as shown in Fig. 4-3:

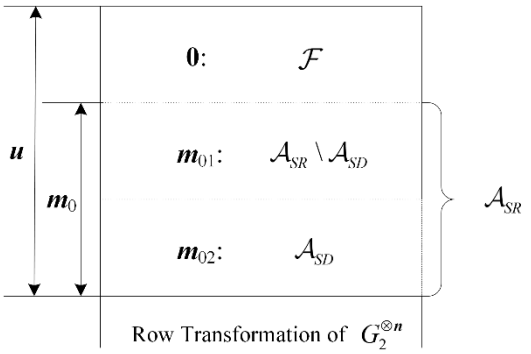


Fig. 4-3 The polar coding construction for BC phase

- $\mathcal{A}_{SR} \setminus \mathcal{A}_{SD}$ represents the information set of the channel W_{SR} and the frozen set of the channel W_{SD}. To interpret it with the idea of channel polarization, it means that $\mathcal{A}_{SR} \setminus \mathcal{A}_{SD}$ contains the bit channels of $I(X_0;Y_1)-I(X_0;Y_0)$ which are "good" in W_{SR} but are "bad" in W_{SD};
- $\mathcal{A}_{SD}$ is the information set of the channels W_{SR} and W_{SD}, i.e., $\mathcal{A}_{SD}$ contains the bit channels of $I(X_0;Y_0)$ which are "good" in W_{SR} and W_{SD};
- $\mathcal{F}=[N_{BC}]\setminus\mathcal{A}_{SR}$ denotes the frozen set of the channels W_{SR} and W_{SD}, i.e., $\mathcal{F}$ contains the bit channels of $1-I(X_0;Y_1)$ which are "bad" in W_{SR} and W_{SD}.

The basic idea of polar coding in the BC phase as shown in Fig. 4-3 is that the source node transmits the information bits belonging to the information set $\mathcal{A}_{SR}$ to the relay node and the destination node, and as $W_{SD} \preceq W_{SR}$ and $\mathcal{A}_{SD} \subseteq \mathcal{A}_{SR}$, the destination node can only receive the information bits belonging to $\mathcal{A}_{SD}$. It should be noted that $\mathcal{A}_{SD} \bigcup (\mathcal{A}_{SR} \setminus \mathcal{A}_{SD}) \bigcup \mathcal{F} = [N_{BC}]$.

The above analysis shows that messages from the source node can directly generate a codeword $\boldsymbol{x}_0$ of block length N_{BC} based on $\mathcal{A}_{SR}$, and then broadcast it to the relay

node and the destination node. According to the structure in Fig. 4-3, the input vector $\boldsymbol{u}$ of length N_{BC} in the source node can be divided into three independent sub-vectors: $\boldsymbol{m}_{01} = \boldsymbol{u}_{\mathcal{A}_{SR}\setminus\mathcal{A}_{SD}}$, $\boldsymbol{m}_{02} = \boldsymbol{u}_{\mathcal{A}_{SD}}$ and $\boldsymbol{u}_{\mathcal{F}}$. Then, the encoder output can be expressed as

$$\boldsymbol{x}_0 = \boldsymbol{m}_{01} G_{\mathcal{A}_{SR}\setminus\mathcal{A}_{SD}} + \boldsymbol{m}_{02} G_{\mathcal{A}_{SD}} + \{0\}^{N_{BC}-|\mathcal{A}_{SR}|} G_{\mathcal{F}}, \tag{4-8}$$

where $G_{\mathcal{A}_{SR}\setminus\mathcal{A}_{SD}}$, $G_{\mathcal{A}_{SD}}$ and $G_{\mathcal{F}}$ are the submatrices of the generator matrix of polar codes consisting of rows with indices in $\mathcal{A}_{SR}\setminus\mathcal{A}_{SD}$, $\mathcal{A}_{SD}$ and $\mathcal{F}$, respectively.

When block length $N_{BC} \to \infty$, the polar coding rate generated is

$$R_0 = \frac{|\mathcal{A}_{SR}|}{N_{BC}}. \tag{4-9}$$

As can be known from Theorem 2-1,

$$\lim_{N_{BC}\to\infty} R_0 = I(X_0;Y_1). \tag{4-10}$$

After receiving $\boldsymbol{y}_1$, the relay node can reliably decode the source message using the SC decoding algorithm given the frozen bits $\boldsymbol{u}_{\mathcal{F}}$. However, as $W_{SD} \preceq W_{SR}$, i.e., $I(X_0;Y_1) > I(X_0;Y_0)$, the destination node cannot reliably reconstruct the source message, but can only store the received $\boldsymbol{m}_0$ and not start decoding until the MAC phase completes. Suppose that the destination node already has the a priori information of $\boldsymbol{m}_{01}$ before decoding (the frozen bit $\boldsymbol{u}_{\mathcal{F}}$ is also known to the polar decoder for the destination node), and then the destination node can also recover the message $\boldsymbol{m}_{02}$. As both the source node and the relay node contain the message vector $\boldsymbol{m}_{01}$, in the MAC phase the relay node and/or the source node will collaborate to assist the transmission of $\boldsymbol{m}_{01}$ in order for the destination node to resolve the uncertainty about the source message $\boldsymbol{m}_0$.

4.3.2 CPMR Transmission Strategy in MAC Phase

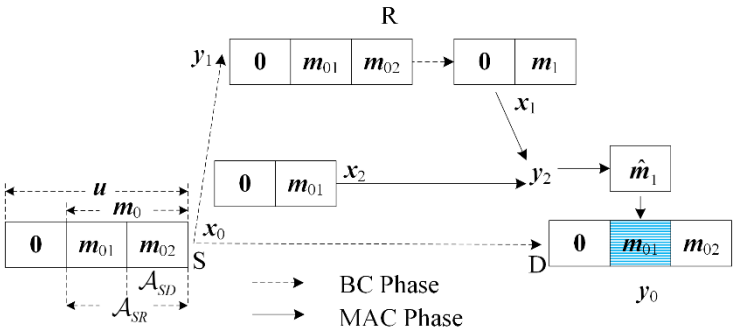
Although the destination node has received $N_{BC}I(X_0;Y_0)$ bits of information, it still needs $N_{BC}(I(X_0;Y_1)-I(X_0;Y_0))$ bits of information (i.e., information of $\boldsymbol{m}_{01}$) so as to reliably decode the source message $\boldsymbol{m}_0$. Although the relay node can recover all of $\boldsymbol{m}_0$, it only needs to transmit part of the bits in $\boldsymbol{m}_0$ to resolve the uncertainty about $\boldsymbol{m}_0$ in the destination node. Define this part of the bit vector as $\boldsymbol{m}_1$ (if the relay node decodes correctly, then $\boldsymbol{m}_1=\boldsymbol{m}_{01}$). Then $\boldsymbol{m}_1$ can be encoded into a codeword $\boldsymbol{x}_1$ of block length N_{MAC} with rate

$$\lim_{N_{MAC}\to\infty} R_1 = \lim_{N_{MAC}\to\infty} \frac{|\mathcal{A}_{SR}|-|\mathcal{A}_{SD}|}{N_{MAC}}$$

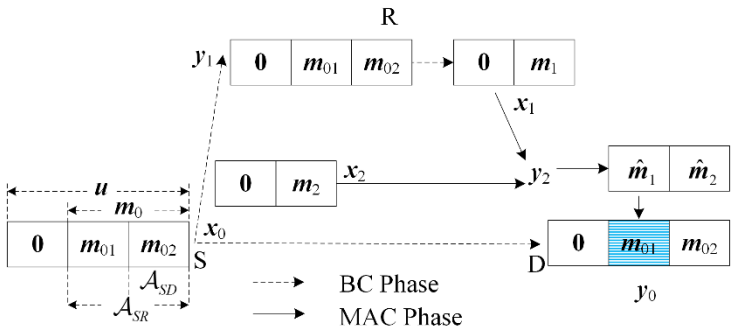
$$= \frac{N_{BC}}{N_{MAC}}(I(X_0;Y_1)-I(X_0;Y_0)) \tag{4-11}$$

Such a transmission strategy is known as the CPMR Protocol.

Meanwhile, the source node also sends a codeword $\boldsymbol{x}_2$ of block length N_{MAC} and there is a certain correlation between $\boldsymbol{x}_2$ and $\boldsymbol{x}_1$ sent by the relay node. That is to say, when the relay node sends $\boldsymbol{x}_1$, the source node sends $r\boldsymbol{x}_1+(1-r)\boldsymbol{x}_2$, where $r\in(0,1)$ is the correlation parameter between codewords $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. However, as pointed out by Ref. [18], the half-duplex relay channel capacity in Equation (4-4) is not sensitive to the correlation parameter *r*, so the coding strategy for the relay node can be simplified so that only the two fundamental extremes of the correlation parameter need to be considered, i.e., $r=0$ and $r=1$. Hence we only consider the two fundamental extremes of the correlation parameter so as to simplify the system design, as illustrated by the two CPMR transmission strategies in Fig. 4-4.



a) r=0



b) r=1

Fig. 4-4 The diagram of the half-duplex relay channel with CPMR protocol

(1) $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ completely independent, i.e., $r = 0$.

The source node sends a new message $\boldsymbol{m}_2$, and encodes it into a codeword $\boldsymbol{x}_2$ with rate

$$\lim_{N_{MAC}\to\infty} R_2 = I(X_2; Y_2 \mid X_1). \tag{4-12}$$

By reviewing section 2.4, we know that the polar code is a channel coding technology capable of achieving the symmetric capacity of B-DMCs. The basic principle of this technology is to convert the repeated use of a specific B-DMC into the individual use of each bit channel after polarization. Polarized bit channels are either “good” or “bad”, which is a phenomenon called channel polarization. Sasoglu et al. extended this phenomenon into the two-user multiple-access channel (T-MAC), and pointed out that the channel polarization phenomenon also exists in T-MAC, but different to single-user polarization with two extremes, it can be polarized into five different possible extremes, and uncoded transmission based on these five extreme bit

channels is optimal [75]. Compared with the single-user phenomenon, the polar coding scheme constructed from the T-MAC polarization phenomenon not only inherits the same ending and decoding complexity but also has the same block error probability. Meanwhile, Sasoglu et al. also proved that, through channel polarization, polar codes can reach any point in the capacity region of T-MAC.

Define the channel $P:\mathcal{X}_1\times\mathcal{X}_2\to\mathcal{Y}_2$ as T-MAC, where the RD and SD channels are B-DMCs respectively, and then R_1 and R_2 satisfy such relations [91]:

$$0\le R_1\le I(X_1;Y_2\mid X_2), \tag{4-13}$$

$$0\le R_2\le I(X_2;Y_2\mid X_1), \tag{4-14}$$

$$0\le R_1+R_2\le I(X_1X_2;Y_2). \tag{4-15}$$

Define $\mathcal{K}(P)=(I(X_1;Y_2\mid X_2),I(X_2;Y_2\mid X_1),I(X_1X_2;Y_2))\in\Re^3$. If

$$M=\{(0,0,0),(0,1,1),(1,0,1),(1,1,1),(1,1,2))\subset\Re^3,$$

and for $p\in\Re^3$, let $d(p,M)=\min_{x\in M}\|p-x\|$ represent the distance from p to M, and then for any $\delta>0$, there is [75]

$$\lim_{N\to\infty}\frac{|\{i\in[N]:d(\mathcal{K}(P^{(i)}),M)\ge\delta\}|}{N}=0, \tag{4-16}$$

where $P^{(i)}$ denotes the ith bit channel after the T-MAC polarization.

According to Equation (4-16), we have five possible extremes as shown in Fig. 4-5 [75]:

- (000) represents the bit channels which do not output any information useful for input decision, i.e., the bit channels observed by the two users are all “bad”;
- (011) and (101) represent the bit channels which provide entirely useful information for deciding the input of one user but do not provide any information useful for deciding the input of the other user, i.e., one user observes the “good” bit channels while the other user observes the “bad” bit

channels;

- (111) denotes the competing bit channels, i.e., if either of the two users transmits at the rate of 0, then the other user observes "good" bit channels;
- (112) represents the bit channels where all the outputs can reliably determine the inputs, i.e., both the two users observe "good" bit channels.

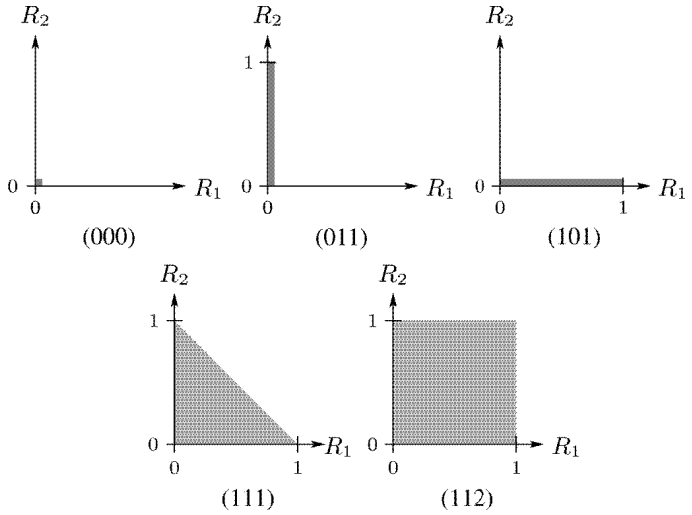


Fig. 4-5 Five limiting regions for T-MAC after polarizing

In the MAC phase, the source and the relay node send the codeword $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ at the same time, and after N independent use of the channel P receive $\boldsymbol{y}_2$. When $N\to\infty$, T-MAC starts polarizing: almost all bit channels $P^{(i)}$ tend to the five extreme bit channels shown in Fig. 4-5. It should be noted that when estimating the symbol of the ith bit channel, assume that the receiver has already received the a priori information of the previous $(i-1)$th estimated symbol $(\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2)_1^{i-1}$, and therefore based on this assumption and the five extreme scenarios in Fig. 4-5, we come up with the following coding schemes.

$$\boldsymbol{x}_1 = \boldsymbol{u}_1 G_{\mathcal{B}_1} + \{0\}^{N_{MAC}-|\mathcal{B}_1|} G_{\mathcal{F}_1}. \tag{4-17}$$

$$\boldsymbol{x}_2 = \boldsymbol{u}_2 G_{\mathcal{B}_2} + \{0\}^{N_{MAC}-|\mathcal{B}_2|} G_{\mathcal{F}_2}. \tag{4-18}$$

where $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are respectively the input vectors of length N_{MAC} in the relay node

and the source node, $\mathcal{B}_1$ and $\mathcal{B}_2$ are respectively the information sets to be sent by the relay node and the destination node, in which, $\mathcal{B}_1 \subset (\boldsymbol{u}_1)_1^{N_{MAC}}$, $\mathcal{B}_2 \subset (\boldsymbol{u}_2)_1^{N_{MAC}}$, $\mathcal{F}_1$ and $\mathcal{F}_2$ are respectively the frozen sets of the relay node and the destination node, i.e., $\mathcal{F}_1 = [N_{MAC}] \setminus \mathcal{B}_1$, $\mathcal{F}_2 = [N_{MAC}] \setminus \mathcal{B}_2$. The values of these two sets are already known to the destination node, and $G_{\mathcal{B}_1}$, $G_{\mathcal{B}_2}$, $G_{\mathcal{F}_1}$ and $G_{\mathcal{F}_2}$ are the submatrices of the generator matrix of polar codes consisting of rows with indices in $\mathcal{B}_1$, $\mathcal{B}_2$ $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively.

For an arbitrarily small $\epsilon > 0$, the information sets $\mathcal{B}_1$ and $\mathcal{B}_2$ can be set respectively according to the following rules [75]:

- If $\| \mathcal{K}(P^{(i)}) - (0,0,0) \| < \epsilon$, then $u_{1,i} \notin \mathcal{B}_1$ and $u_{2,i} \notin \mathcal{B}_2$;
- If $\| \mathcal{K}(P^{(i)}) - (0,1,1) \| < \epsilon$, then $u_{1,i} \notin \mathcal{B}_1$ and $u_{2,i} \in \mathcal{B}_2$;
- If $\| \mathcal{K}(P^{(i)}) - (1,0,1) \| < \epsilon$, then $u_{1,i} \in \mathcal{B}_1$ and $u_{2,i} \notin \mathcal{B}_2$, or $u_{1,i} \notin \mathcal{B}_1$ and $u_{2,i} \in \mathcal{B}_2$;
- If $\| \mathcal{K}(P^{(i)}) - (1,1,1) \| < \epsilon$, then $u_{1,i} \in \mathcal{B}_1$ and $u_{2,i} \in \mathcal{B}_2$;
- If $\| \mathcal{K}(P^{(i)}) - (1,1,2) \| < \epsilon$, then $u_{1,i} \notin \mathcal{B}_1$ and $u_{2,i} \notin \mathcal{B}_2$;
- Otherwise, then $u_{1,i} \notin \mathcal{B}_1$ and $u_{2,i} \notin \mathcal{B}_2$.

The above rules for selecting $\mathcal{B}_1$ and $\mathcal{B}_2$ ensures that information bits can be passed through "good" bit channels provided that the previous symbol decoding is assumed to be correct. Therefore, after the destination node receives $\boldsymbol{y}_2$, it can attempt to carry out successive decoding on $\boldsymbol{y}_2$ in the order of $(\boldsymbol{u}_{1,1}, \boldsymbol{u}_{2,1}), (\boldsymbol{u}_{1,2}, \boldsymbol{u}_{2,2}), \cdots$, so as to obtain a relatively low block error probability.

(2) $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are completely correlated, i.e., $r = 1$.

The source node and the relay node also send $\boldsymbol{m}_{01}$ through the same codeword $\boldsymbol{x}_1$

and together help the destination node resolve the uncertainty about the source message $\boldsymbol{m}_0$.

After the decoding completes in the MAC phase, the destination node can use the estimated partial message $\hat{\boldsymbol{m}}_1$ as a priori information to decode $\boldsymbol{y}_0$ received in the BC phase with the aim of recovering the source message $\boldsymbol{m}_0$. At this point, in order to decode $\boldsymbol{y}_0$, the destination node will treat it as decoding the polar codes constructed from the received information set $\mathcal{A}_{SD}$. As the destination node has already known the a priori information $\hat{\boldsymbol{m}}_1$, its corresponding rate will become smaller, and then there is

$$\lim_{N_{BC}\to\infty} R'_0 = \lim_{N_{BC}\to\infty} \frac{|\mathcal{A}_{SD}|}{N_{BC}} = I(X_0;Y_0). \tag{4-19}$$

From the above design of the CPMR protocol, we know that when $r = 0$, the overall transmission rate R_{all} is

$$R_{all} = tR_0 + (1-t)R_2. \tag{4-20}$$

When N_{BC} and N_{MAC} tend to infinity, from Equation (4-10) and Equation (4-12) we get

$$\lim_{N_{BC},N_{MAC}\to\infty} R_{all} = tI(X_0;Y_1) + (1-t)I(X_2;Y_2 \mid X_1). \tag{4-21}$$

The above equation is exactly the half-duplex relay channel capacity in Equation (4-4). As the optimization of *t* has already made the two terms on the right-hand side of the equality equal to each other in Equation (4-4), we will just focus on the first term.

Similarly, when $r = 1$, there is

$$R_{all} = tR_0. \tag{4-22}$$

In this case, $I(X_2;Y_2 \mid X_1) = 0$, so we have

$$\lim_{N_{BC},N_{MAC}\to\infty} R_{all} = tI(X_0;Y_1). \tag{4-23}$$

The above equation is exactly the half-duplex relay channel capacity in Equation (4-4) when $r=1$.

To sum up, when $r=0$ and $r=1$, with the CPMR protocol as the transmission strategy, information rates can all asymptotically achieve the half-duplex relay channel capacity of Equation (4-4).

4.3.3 Analysis of the Asymptotic Performance

The following paragraphs discuss in great detail the asymptotic performance of the CPMR protocol based on polar codes when $r=0$ and $r=1$.

Define the block error probability of the CPMR scheme as follows:

$$P_e=\begin{cases}\Pr\{(\hat{\boldsymbol{m}}_0,\hat{\boldsymbol{m}}_2)\neq(\boldsymbol{m}_0,\boldsymbol{m}_2)\}, & r=0\\ \Pr\{\hat{\boldsymbol{m}}_0\neq\boldsymbol{m}_0\}, & r=1\end{cases}. \tag{4-24}$$

Then, the following theorem gives the upper bound on the block error probability P_e.

Theorem 4-1 (Asymptotic Performance of the CPMR Scheme): When $r=0$ and $r=1$, for an arbitrary $0<\beta<1/2$, the block error probability of the CPMR scheme based on the SC decoding algorithm is upper bounded by

$$P_e\le\max\{O(2^{-(N_{BC})^{\beta}}),O(2^{-(N_{MAC})^{\beta}})\}, \tag{4-25}$$

where the block length N_{BC} and N_{MAC} tend to infinity.

Proof: Let $\mathcal{E}$ represent the error event $\{(\hat{\boldsymbol{m}}_0,\hat{\boldsymbol{m}}_2)\neq(\boldsymbol{m}_0,\boldsymbol{m}_2)\}$ existing at the destination node when $r=0$ or the error event $\{\hat{\boldsymbol{m}}_0\neq\boldsymbol{m}_0\}$ existing at the destination node when $r=1$. Let $\mathcal{E}_{BC}$ be the error event $\{\hat{\boldsymbol{m}}_0\neq\boldsymbol{m}_0\}$ when the relay decodes the received source message $\boldsymbol{m}_0$ at the end of BC phase, and define $\mathcal{E}_{MAC}$ respectively as the error event $\{(\hat{\boldsymbol{m}}_1,\hat{\boldsymbol{m}}_2)\neq(\boldsymbol{m}_1,\boldsymbol{m}_2)\}$ under the condition of $r=0$ or the error event $\{\hat{\boldsymbol{m}}_1\neq\boldsymbol{m}_1\}$ under the condition of $r=1$ existing for the destination node when the MAC phase ends. Then, $\mathcal{E}_{BC}^{c}$ and $\mathcal{E}_{MAC}^{c}$ respectively represent the complementary

events of the event $\mathcal{E}_{BC}$ and the event $\mathcal{E}_{MAC}$. Accordingly, we have

$$\begin{aligned} P_e &= \Pr(\mathcal{E}) \\ &= \Pr(\mathcal{E}_{BC})\Pr(\mathcal{E} \mid \mathcal{E}_{BC}) + \Pr(\mathcal{E}_{BC}^c)\Pr(\mathcal{E} \mid \mathcal{E}_{BC}^c) \\ &\le \Pr(\mathcal{E}_{BC}) + \Pr(\mathcal{E} \mid \mathcal{E}_{BC}^c). \end{aligned} \tag{4-26}$$

Based on the SC decoding performance of single-user polar codes demonstrated in Theorem 2-1, we know that after the relay node decodes the source message $\boldsymbol{m}_0$ transmitted via the SR link, the block error probability obtained is

$$\Pr(\mathcal{E}_{BC}) \le O(2^{-(N_{BC})^{\beta}}). \tag{4-27}$$

The second term in Equation (4-26) can be further rewritten into

$$\begin{aligned} \Pr(\mathcal{E} \mid \mathcal{E}_{BC}^c) &= \Pr(\mathcal{E}_{MAC} \mid \mathcal{E}_{BC}^c)\Pr(\mathcal{E} \mid \mathcal{E}_{BC}^c, \mathcal{E}_{MAC}) \\ &\quad + \Pr(\mathcal{E}_{MAC}^c \mid \mathcal{E}_{BC}^c)\Pr(\mathcal{E} \mid \mathcal{E}_{BC}^c, \mathcal{E}_{MAC}^c) \\ &\le \Pr(\mathcal{E}_{MAC} \mid \mathcal{E}_{BC}^c) + \Pr(\mathcal{E} \mid \mathcal{E}_{BC}^c, \mathcal{E}_{MAC}^c). \end{aligned} \tag{4-28}$$

Then Equation (4-28) is computed for $r = 0$ and $r = 1$.

(1) $r = 0$

The first term in Equation (4-28) represents the error probability of the event $\{(\hat{\boldsymbol{m}}_1, \hat{\boldsymbol{m}}_2) \neq (\boldsymbol{m}_1, \boldsymbol{m}_2)\}$ on the condition that the source message received by the relay node is correctly decoded after the BC phase.

Let $(\hat{u}_{1,i}, \hat{u}_{2,i}) = \phi_i(\boldsymbol{y}_2, (\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2)_1^{i-1})$ represent the decoding rule for estimating $(u_{1,i}, u_{2,i})$ when $(\boldsymbol{y}_2, (\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2)_1^{i-1})$ is given. This corresponds to Arikan's genie-aided decision rule [63], where genie firstly provides the information $(\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2)_1^{i-1}$ previously correctly decoded and then the decoder performs estimation on the current $(u_{1,i}, u_{2,i})$ based on the received $\boldsymbol{y}_2$. The existence of such an assumption is reasonable because the decoder already knows the value of each frozen bit and its corresponding index,

and thus it can carry out successive decoding in the order of $(\boldsymbol{u}_{1,1},\boldsymbol{u}_{2,1}),(\boldsymbol{u}_{1,2},\boldsymbol{u}_{2,2}),\ldots$ based on the SC decoding algorithm.

From the discussion on T-MAC capacity region by Ref. [75], we can get the following conclusion.

For T-MAC in the MAC phase, there are two polar code sequences of sufficiently-long block lengths, whose rate pairs satisfy

$$\begin{cases} R_1 > I(X_1;Y_2)-\epsilon \\ R_2 > I(X_2;Y_2 \mid X_1)-\epsilon \end{cases}. \tag{4-29}$$

Then, the upper bound on the average block error probability for decoding $\boldsymbol{y}_2$ received at the destination node is

$$\Pr(\mathcal{E}_{MAC} \mid \mathcal{E}_{BC}^{c}) \le O(2^{-(N_{MAC})^{\beta}}), \tag{4-30}$$

where, $\epsilon > 0$, and $0 < \beta < 1/2$.

(2) $r = 1$

The first term in Equation (4-28) represents the error probability of the event $\{\hat{\boldsymbol{m}}_1 \neq \boldsymbol{m}_1\}$ when the received source message is correctly decoded by the relay node in the BC phase. At this point, according to the rate design of the CPMR scheme hereinbefore, the source and relay nodes send exactly the same codeword at rates smaller than the capacity of the transmission channel, so we can get

$$\Pr(\mathcal{E}_{MAC} \mid \mathcal{E}_{BC}^{c}) \le O(2^{-(N_{MAC})^{\beta}}). \tag{4-31}$$

The second term in Equation (4-28) represents the error probability of recovering the source message $\boldsymbol{m}_0$ received in the BC phase after the successful decoding of the partial message $\boldsymbol{m}_1$. Therefore, for any given rate $R'_0 < I(X_0;Y_0)$,

$$\Pr(\mathcal{E} \mid \mathcal{E}_{BC}^{c}, \mathcal{E}_{MAC}^{c}) \le O(2^{-(N_{BC})^{\beta}}). \tag{4-32}$$

Combine Equation (4-30), Equation (4-31) and Equation (4-32), and then we get

that Equation (4-28) is upper bounded by

$$\Pr(\mathcal{E} \mid \mathcal{E}_{BC}^{c}) \le \max\{O(2^{-(N_{BC})^{\beta}}), O(2^{-(N_{MAC})^{\beta}}) . \tag{4-33}$$

When the block length N_{BC} and N_{MAC} tend to infinity, we combine Equation (4-27) and Equation (4-33) and substitute it into Equation (4-26), and then Equation (4-25) in this theorem holds. □

The preceding analysis shows that only by satisfying that W_{SD} is the stochastically degraded channel of the channel W_{SR} can Theorem 4-1 be proved. As we all know, generally speaking, the DF relay protocol can only reach the capacity of physically degraded half-duplex relay channels [77, 135], and if a half-duplex relay channel is stochastically degraded, its channel capacity will be consistent with the DF achievable rate. However, physical degradation implies stochastic degradation, so the CPMR scheme proposed herein also applies to physically degraded half-duplex relay channels.

From the above analysis of the asymptotic performance of the CPMR scheme, we know that the proposed scheme can reach the capacity of physically degraded half-duplex relay channels when the block length is infinite. Compared with the case of conventional LDPC coding schemes, the above conclusion is obtained by just using polar encoding and decoding methods of low constructive complexity, so the complexity of the encoder and decoder in the relay channel is only $O(N \log N)$. In addition, if all links in the relay channel are BEC channels, then the construction complexity of polar codes is only $O(N)$ [63].

4.4 CPMR Transmission Scheme of Finite Block Lengths

This section discusses in detail how to realize the CPMR transmission scheme based on polar codes of finite block lengths. In order to facilitate the design and analysis of

system models, from the perspective of practical application of wireless communications, we build a system model for the half-duplex AWGN relay channel with BPSK modulation. Assume that the channel state information is given for the three nodes in the relay channel. This section not only discusses the influence of the time allocation factor *t* on system capacity, but also designs the coding method and the receiver structure based on the channel model and lastly it provides the simulated system performance as well as its analysis.

4.4.1 System Model

Fig. 4-6 illustrates a practical relay system model, where, with the help of the wireless access point (R), the mobile user terminal (S) sends data to the base station (D). The relay system transmits data in time-division half-duplex mode. This section also normalizes the system time, and defines *t* as the system time allocation factor. During the first time interval *t* (BC phase), S sends data to R and D at the same time; during the second time interval $(1-t)$ (MAC phase), R and S send data to D at the same time while R at this point does not receive any data.

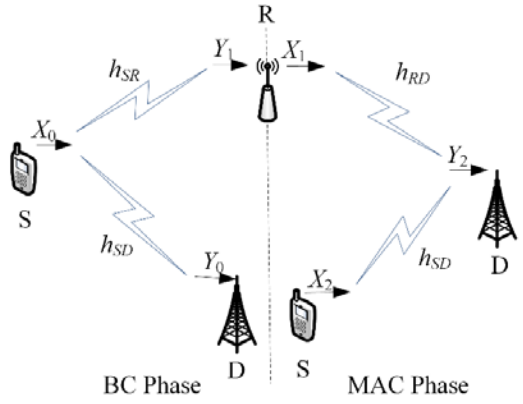


Fig. 4-6 A half-duplex single-relay system model

As this section only focuses on analyzing the actual performance of the CPMR scheme, the system model is simplified as shown in Fig. 4-7. Assume the wireless relay access point R is located on the connector between S and D and it normalizes the

distance between S and D to 1. Use $d \in (0,1)$ to represent the distance from S to R, then the distance from R to D is $(1-d)$. In this study, only large-scale path loss is considered, so the channel parameters of the SR, SD and RD channels in the system respectively are

$$\begin{cases} h_{SR} = \sqrt{1/d^{\alpha}} \\ h_{SD} = 1 \\ h_{RD} = \sqrt{1/(1-d)^{\alpha}} \end{cases}, \tag{4-34}$$

where α denotes the fading coefficient of the channel, assuming $\alpha = 2$. Also, assume all the noise in the system is white Gaussian noise with mean of 0 and variance of 1.

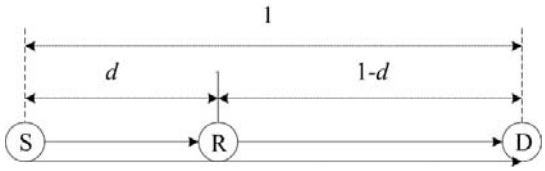


Fig. 4-7 The simplified model of the half-duplex single-relay system

Based on the above definitions and assumptions, we have the following channel models:

$$Y_1[i] = \sqrt{P_S} h_{SR} X_0[i] + Z_R[i] \qquad i = 1, \ldots, N_{BC}, \tag{4-35}$$

$$Y_0[i] = \sqrt{P_S} h_{SD} X_0[i] + Z_D[i] \qquad i = 1, \ldots, N_{BC}, \tag{4-36}$$

$$Y_2[i] = \sqrt{P'_S} h_{SD} X_2[i] + \sqrt{P_R} h_{RD} X_1[i] + Z'_D[i] \qquad i = 1, \ldots, N_{MAC}, \tag{4-37}$$

where $Z_R[i]$, $Z_D[i]$ and $Z'_D[i]$ are respectively the additive white Gaussian noises of the SR and SD links in the BC phase and the additive white Gaussian noise of the SD link in the MAC phase, P_S denotes the source transmission power in the BC phase, and $P_S = E[X_0[i]^2]$, P'_S is the source transmission power in the MAC phase, and $P'_S = E[X_2[i]^2]$ and P_R denotes the relay transmission power in the MAC phase, and $P_R = E[X_1[i]^2]$.

Assume an average global transmission power is constrained by [18]:

$$\Theta: tP_S + (1-t)(P'_S + P_R) \le P, \tag{4-38}$$

where P denotes the total transmission power of the relay system. As the noise power is normalized to 1, P is equivalent to the overall SNR of the system. Here we assume that the source node and the destination node have an overall average power limit rather than limit their transmission power separately because by flexibly allocating the transmission power between S and R, a higher system capacity can be achieved.

As previously mentioned, in the MAC phase, the relay node and the source node can send partial messages to the destination node (i.e., $r = 0,1$) in two cooperative modes to address the uncertainty about the source message received by the destination node in the BC phase. The only difference between the cooperative mode when $r = 1$ and the mode when $r = 0$ is that there does not exist interference between signals in the MAC phase so the cooperative mode of $r = 1$ can simply be regarded as an end-to-end polar coding scheme, which simultaneously decodes the partial message from the source node and the destination node. Therefore, next we only consider the relatively more complex case of $r = 0$ and evaluate the system performance through simulation.

4.4.2 Analysis of Time Fraction

The CPMR scheme proposed herein is based on Equation (4-4) whose channel capacity is our goal to be achieved by polar coding designs. When the power constrains of S and R are

$$\begin{cases} P'_S \le P \\ P_R \le P \end{cases}, \tag{4-39}$$

it can be proved that when the two terms in the min(.) function of Equation (4-4) are equal, the capacity C of the half-duplex relay channel is maximum [18]. Hence, when the power constrains in Equation (4-39) are satisfied, the numerical analysis method

can be used to obtain the optimal power allocation between P'_S and P_R as well as the optimal time allocation factor *t*.

The analysis of the factors influencing the channel capacity is crucial to the design of channel coding schemes because it provides the basic capacity limit for reliable communication. Fig. 4-8 demonstrates the capacity comparison between the direct transmission channel and the half-duplex relay channel with *t* valued at the optimal and 0.5 respectively at different SNRs and *d*s when $r = 0$. For a fair comparison, we make the sum of the transmission powers of S and R equal to the transmission power for the direct transmission. By analyzing Fig. 4-8, it is easy get the following three conclusions:

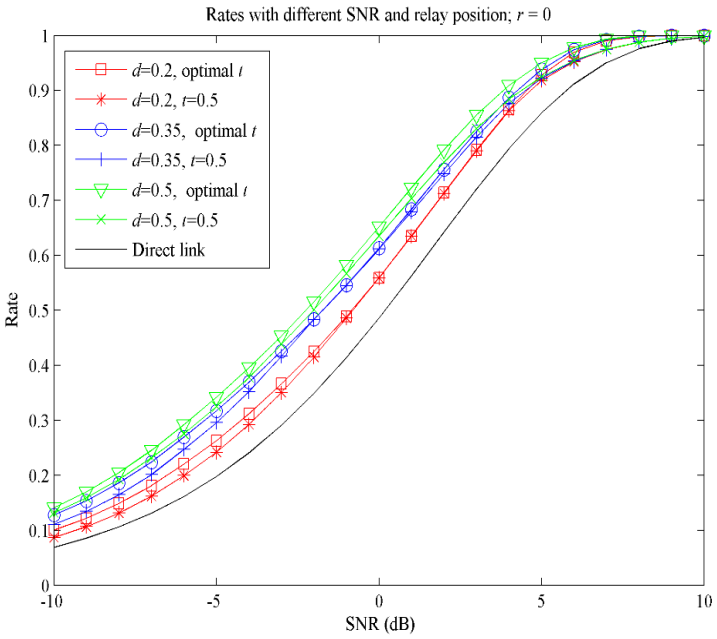


Fig. 4-8 Information rates vs. SNR for different relay positions on an AWGN relay channel with BPSK modulation

1) by means of forwarding information from the relay node, a channel capacity bigger than that of the direct transmission can be obtained, and the closer the relay node is to the source node, the more obvious the gain in relay the system gets;

2) in comparison with the direct transmission, the relay system used at low SNRs can produce bigger gain. However, as SNR increases, the advantage of using relay

over the direct transmission diminishes until there is no advantage at all;

3) compared with the optimal *t*, when choosing $t = 0.5$, the system capacity only sustains a small loss, so the time allocation factor has a relatively small influence on relay channel capacity.

Inspired by the above observations, in order to simplify the system design and guarantee a minimum loss of channel capacity, we used $t = 0.5$ for the system polar coding design during the simulation of the actual half-duplex relay channel model, which also ensures a near-optimal system performance with BPSK modulation at low SNRs.

4.4.3 Construction of Polar Codes in the BI-AWGN Channel

Arikan pointed out that the construction of polar codes is related to specific channels and if one wants to achieve the symmetric capacity other than the BEC channel, it must be done at the cost of greater coding complexity [111]. At this stage, research on polar codes mainly focuses on end-to-end channels while there is still no practical and specific coding scheme for the BI-AWGN channel with DF cooperative relay. We use a coding construction method similar to coding in the BEC channel, which can achieve a good performance in the BI-AWGN channel while maintaining relatively low coding complexity.

In the process of constructing polar codes, assume that in the BC phase the SR channel and the SD channel are respectively BI-AWGN channels with channel capacity $I(X_0;Y_1)$ and $I(X_0;Y_0)$. If the coding construction for the BEC corresponds to that for the BI-AWGN channel, then the "erasure probabilities" of the SR and SD channels can respectively be expressed as

$$\varepsilon_0 = 1 - I(X_0;Y_1), \tag{4-40}$$

$$\varepsilon'_0 = 1 - I(X_0;Y_0). \tag{4-41}$$

Then we use recursive formulae Equation (2-48) and Equation (2-49) to compute the Bhattacharyya parameters, where $Z(W_{SR}) = \varepsilon_0$ and $Z(W_{SD}) = \varepsilon'_0$. Similarly, the "erasure probabilities" corresponding to the SD and RD channels in the MAC phase can be obtained respectively.

Each mutual information in Equation (4-4), which is the expression for the capacity of physically degraded half-duplex relay systems, can be obtained respectively through the following equations

$$I(X_0;Y_1) = H(Y_1) - H(Z), \tag{4-42}$$

$$I(X_0;Y_0) = H(Y_0) - H(Z), \tag{4-43}$$

$$I(X_2;Y_2 \mid X_1) = \frac{1}{2}\left\{ \sum_{b=-1,+1} H(Y_2 \mid X_1 = b) \right\} - H(Z), \tag{4-44}$$

$$I(X_1, X_2;Y_2) = H(Y_2) - H(Z). \tag{4-45}$$

It should be noted that their probability mass functions respectively are

$$f_{Y_1}(y_1) = \sum_{a=-1,+1} f_{X_0}(a) f_Z(y_1 - \sqrt{(h_{SR})^2 P_S}), \tag{4-46}$$

$$f_{Y_0}(y_0) = \sum_{a=-1,+1} f_{X_0}(a) f_Z(y_0 - \sqrt{(h_{SD})^2 P_S}), \tag{4-47}$$

$$f_{Y_2|X_1}(y_2 \mid b) = \sum_{a=-1,+1} f_{X_2|X_1}(a \mid b) f_Z(y_2 - a\sqrt{(h_{SD})^2 P'_S} - b\sqrt{(h_{RD})^2 P_R}), \tag{4-48}$$

$$f_{Y_2}(y_2) = \sum_{a=-1,+1} \sum_{b=-1,+1} f_{X_1,X_2}(a,b) f_Z(y_2 - a\sqrt{(h_{SD})^2 P'_S} - b\sqrt{(h_{RD})^2 P_R}), \tag{4-49}$$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp(\frac{-z^2}{2}), \tag{4-50}$$

where $f_{X_0}(-1) = f_{X_0}(+1) = \frac{1}{2}$, and the optimal distribution of $f_{X_1,X_2}(a,b)$ is

$$f_{X_1,X_2}(1,1) = f_{X_1,X_2}(-1,-1) = \frac{1}{4},$$

$$f_{X_1,X_2}(1,-1) = f_{X_1,X_2}(-1,1) = \frac{1}{4}. \tag{4-51}$$

4.4.4 Receiver Structure

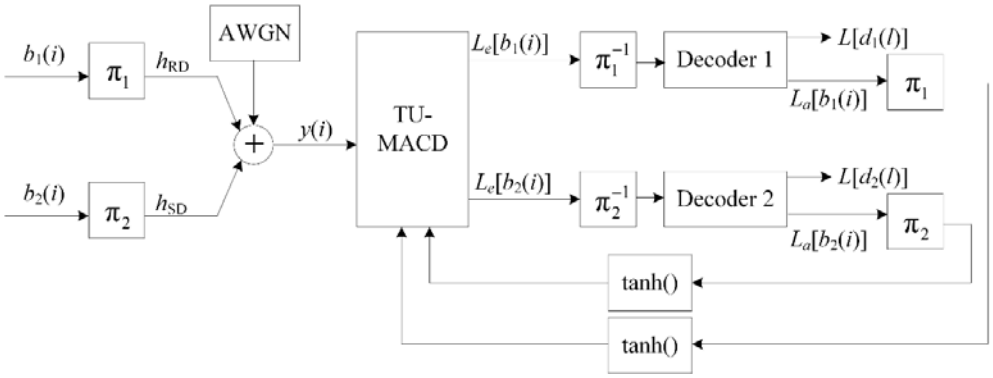


Fig. 4-9 Block diagram of the receiver design for polar coded relay system

From the preceding discussions, we know that when $r = 0$, the difficulty in the system design lies in how to correctly recover the partial message individually sent by the relay node while there is interference from new messages sent from the source. Under such circumstances, the codewords $\boldsymbol{x}_0$, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ respectively received by the destination node in the BC phase and the MAC phase construct distributed polar codes. At this point, the destination node can perform joint decoding on the received sequence in order to obtain the source message. Although the T-MAC polarization method discussed in the previous section can be used to perform joint detection on the received sequence in order to get the optimal performance, but due to high complexity of such joint detection, it is impractical to carry it out, which is why we use a iterative processing method (i.e., the "Turbo principle") to address the issue of multiuser detection and decoding and why we designed the JISPIC receiver structure for the polar codes. The structure shown in Fig. 4-9 consists of two basic components - a two-user multiple-access channel detector (TU-MACD) and two polar decoders. These two components are separated by the interleavers π_1, π_2 and the deinterleavers π_1^{-1}, π_2^{-1}. TU-MACD is composed of a soft parallel interference cancellation (SPIC) part and

single-user LLR calculation (SU-LLRC) parts. In the following paragraphs, each component of the JISPIC decoder is discussed in detail.

(1) TU-MACD

Firstly we consider a polar encoder of block length N_S (because $t = 0.5$, for the convenience of expression, assume hereinafter $N_{BC} = N_{MAC} = N_S$) with rate R_k for $k = 1, 2$. For $1 \le l \le R_k N_S$, the input to the encoder is $\boldsymbol{d}_k = (d_k(1), \ldots, d_k(l), \ldots d_k(R_k N_S))$, then for $1 \le i \le N_S$, the corresponding output to the encoder is $\boldsymbol{b}_k = (b_k(1), \ldots, b_k(i), \ldots b_k(N_S))$. Before the source node and the relay node transmit messages, the messages will respectively be passed through a pair of interleavers π_1 and π_2 with a view to reducing the impact of burst errors input to each polar decoder. In TU-MACD, we employ a soft-input soft-output (SISO) maximum a posteriori probability (MAP) multi-user detector, whose objectives are to compute the soft information of the symbol bits transmitted from the relay node and the source node according to the received signal $\boldsymbol{y} = (y(1), \ldots, y(i), \ldots y(N_S))$ at the destination node and computer the a priori information of each symbol bit. For $b_1(i)$, the LLR of the "1" and "-1" transmitted by the SISO detector is

$$L[b_1(i)] = \log \frac{\Pr(b_1(i) = +1 \mid y(i))}{\Pr(b_1(i) = -1 \mid y(i))}. \tag{4-52}$$

Using Bayes' formula, Equation (4-52) can be written as

$$L[b_1(i)] = \log \frac{\sum_{b_1(i)=+1, b_2(i)=\pm 1} \Pr(y_i \mid b_1(i), b_2(i))}{\sum_{b_1(i)=-1, b_2(i)=\pm 1} \Pr(y_i \mid b_1(i), b_2(i))} + \log \frac{\Pr(b_1(i) = +1)}{\Pr(b_1(i) = -1)} \tag{4-53}$$

$$= L_e[b_1(i)] + L_a[b_1(i)], \tag{4-54}$$

where $L_a[b_1(i)]$ denotes the a priori LLR information of $b_1(i)$, which is computed from the previous iteration of the first MAP polar decoder. After being interleaved,

the value is fed back to the TU-MACD. $L_e[b_1(i)]$ is the output extrinsic information delivered by the TU-MACD, which is then de-interleaved and fed into the first channel decoder as the a priori information for the next iteration. Similarly, we can compute the soft information $L[b_2(i)]$, the a priori information $L_a[b_2(i)]$ and the extrinsic information $L_e[b_2(i)]$ for $b_2(i)$.

(2) SPIC

In each iteration, SPIC performs a soft interference cancellation by subtracting the soft estimates of $\boldsymbol{b}_k$ obtained from the previous iteration from the received signal $\boldsymbol{y}$. Suppose that $\hat{\boldsymbol{b}}_k$ is the soft estimates of $\boldsymbol{b}_k$. Then the output of SPIC for $b_1(i)$ can be written as:

$$\tilde{b}_1(i) = y(i) - \sqrt{P'_S}\, h_{SD} \hat{b}_2(i) \tag{4-55}$$

$$= \sqrt{P_R}\, h_{RD} b_1(i) + \sqrt{P'_S}\, h_{SD} (b_2(i) - \hat{b}_2(i)) + z(i), \tag{4-56}$$

where $z(i)$ denotes the white Gaussian noise and the white Gaussian noise vector is $\boldsymbol{z} = (z(1), \ldots, z(N_S))$. $b_2(i) - \hat{b}_2(i)$ is the remaining interference information of the ith codeword, which shows that the soft estimates of code bits are not entirely identical to code bits. $\hat{b}_2(i)$ represents the soft estimates of $b_2(i)$ obtained from the previous iteration. $\hat{b}_2(i)$ can be computed from the following equation [136, 137]

$$\hat{b}_2(i) = 1 \times \Pr(b_2(i) = +1 \mid y(i)) + (-1) \times \Pr(b_2(i) = -1 \mid y(i))$$

$$= \frac{e^{L_e[b_2(i)]} - 1}{e^{L_e[b_2(i)]} + 1} = \tanh\left(\frac{1}{2} L_e[b_2(i)]\right). \tag{4-57}$$

After passing through the SPIC, the multiple-access interference is almost completely eliminated. Similarly, $\hat{b}_1(i)$ and $\tilde{b}_2(i)$ can also be computed. For the first iteration, suppose that the code bits $\boldsymbol{b}_k$ have equal probability (i.e., there is no

corresponding a priori information), and then we have $L_a[b_k(i)]=0$.

(3) SU-LLRC

As the code bits $\boldsymbol{b}_1$ has already obtained the a priori distribution from the extrinsic information delivered by the iterative decoder previously, the probability mass function of $\tilde{b}_1(i)$ can be computed when $b_1(i)$ is known. In the following iterations, SU-LLRC firstly computes the input extrinsic information of the code bits in the form of $L_e[b_1(i)]$, which is given as

$$L_e[b_1(i)]=\log\frac{\Pr(\tilde{b}_1(i)\,|\,b_1(i)=+1)}{\Pr(\tilde{b}_1(i)\,|\,b_1(i)=-1)}$$

$$=\frac{2\sqrt{P_R}h_{RD}}{(\sigma_1(i))^2}\tilde{b}_1(i)\,, \tag{4-58}$$

where the variance for the first decoder is given by the expectation, i.e.,

$$(\sigma_1(i))^2=E\{(\tilde{b}_1(i)-\sqrt{P_R}h_{RD}\hat{b}_1(i)\}\,, \tag{4-59}$$

where $E\{.\}$ is the expectation operation. The SU-LLRC then delivers $L_e[b_1(i)]$ as the a priori information to the decoder. The extrinsic information will become more and more reliable after exchanging the soft estimates and the interference will be suppressed over several iterations. After the last iteration, the MAP decoder computes the a posteriori LLR information $L[d_1(l)]$ for each information bit, i.e., $\hat{d}_1(l)=\mathrm{sgn}(L(d_1(l)))$. Similarly, we can also get $\hat{d}_2(l)$. It should be noted that at the start of the first iteration, the statistics of $L_e[b_1(i)]$ and $L_a[b_1(i)]$ are carried out independently, but in the following iterations, due to the use of common information, they will become increasingly correlated so that the performance will not be further improved through iterations.

4.4.5 Simulation Results and Analysis

In this section, we verify the system performance of the CPMR scheme based on polar codes of finite block lengths. From the conclusion derived from the analysis of Fig. 4-8, we know that the system can obtain bigger relaying gain at low SNRs with BPSK modulation and the use of identical time allocation parameters has a relatively small impact on the relay channel capacity. Therefore, we first determine the overall transmission power to be P = -6 dB, and then compute the optimal power allocation and transmission rate when $t = 0.5$, $d = 0.5$ and $r = 0$ so as to achieve the maximum system channel capacity C, with the simulation parameters as shown in Table 4-1. For better practical performance, the BP decoding is used as the decoding algorithm, and the maximum number of decoding iterations is 200 times. N_{BC} and N_{MAC} are respectively 2^{15}.

Table 4-1 Simulation parameters

P(dB)	P'_S (dB)	P_R (dB)	R_0 (b/s)	R'_0 (b/s)	R_1 (b/s)	R_2 (b/s)	C(b/s)
-6	-10.401	-7.959	0.488	0.162	0.326	0.063	0.275

Although the source node broadcasts the codeword $\boldsymbol{x}_0$ in the BC phase, the destination node cannot decode $\boldsymbol{x}_0$ due to lack of reliable information. Therefore, the decoding performance of the relay node, which serves as the intermediate point for the partial reliable message, determines the reliability of decoding the received $\boldsymbol{x}_0$ at the destination node. If the relay code decodes the codeword $\boldsymbol{x}_0$ correctly, then the destination node's performance in decoding the codeword $\boldsymbol{x}_1$ using the JISPIC receiver in the MAC phase plays a determining role in reconstructing the source message received by the destination node in the BC phase. Fig. 4-10 illustrates the BER performance of decoding $\boldsymbol{x}_1$ with the proposed JISPIC receiver. As the figure

shows, as the number of iterations between TU-MACD and the decoder increases, the decoding performance significantly improves, but when the number of iterations exceeds 2 times, the iterations cease to further improve the performance. This is because after a certain number of iterations, the extrinsic information of the two decoders will become increasingly correlated, which leads to less and less performance improvement.

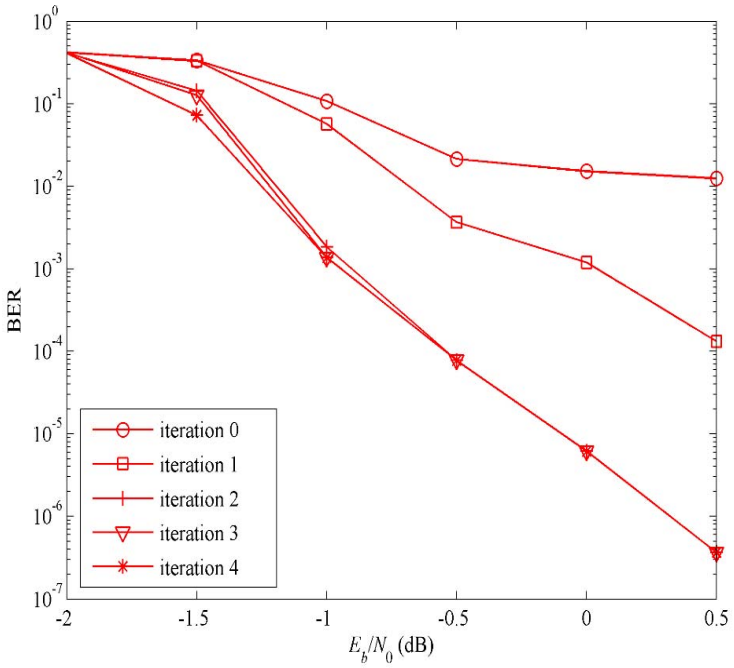


Fig. 4-10 The performance evaluations of the JISPIC receiver for decoding $\boldsymbol{x}_1$

Theorem 4-1 proved that the CPMR scheme for the half-duplex relay system proposed herein is theoretically optimal, and as long as the block length is sufficiently long, the system channel capacity can be achieved. However, as can be seen from the simulation results in Fig. 4-11, the decoding performance of the CPMR scheme of finite block lengths using the BP algorithm has a slow convergence speed, which is a common issue for constructing polar codes of finite block lengths [63]. To further enhance the system performance, the BP decoder in the CPMR scheme can be replaced with the SCLD decoder for decoding polar codes (see Ref. [71, 72] regarding the SCLD decoding algorithm). As can be seen from Fig. 4-11, compared with the CPMR

scheme using the BP decoder, the CPMR scheme which has switched to the SCLD decoder has notably improved its decoding performance. Apart from demonstrating the influence of these two different types of decoders on the CPMR scheme, the diagram also gives the performance of the half-duplex relay system employing the conventional LDPC coding scheme, where LDPC codes are optimized using density evolution together with Gaussian approximation [18]. The simulation results in Fig. 4-11 show that when BER=10^{-5} is used as the standard for measuring reliable communication, the CPMR scheme using the SCLD decoder has a performance about 0.1dB superior to that using the LDPC code. While the time complexity of the BP decoder is $O(N\log N)$, the time coding complexity of the SCLD decoder is $O(L\times N\log N)$ (where, L is the list number used by the SCLD decoder), but when the SCLD decoder is used for decoding, its space complexity is only $O(N)$. Compared with the application of the conventional LDPC coding scheme in the relay system, the CPMR scheme based on polar codes also enjoys relatively low coding complexity, which is only $O(N\log N)$.

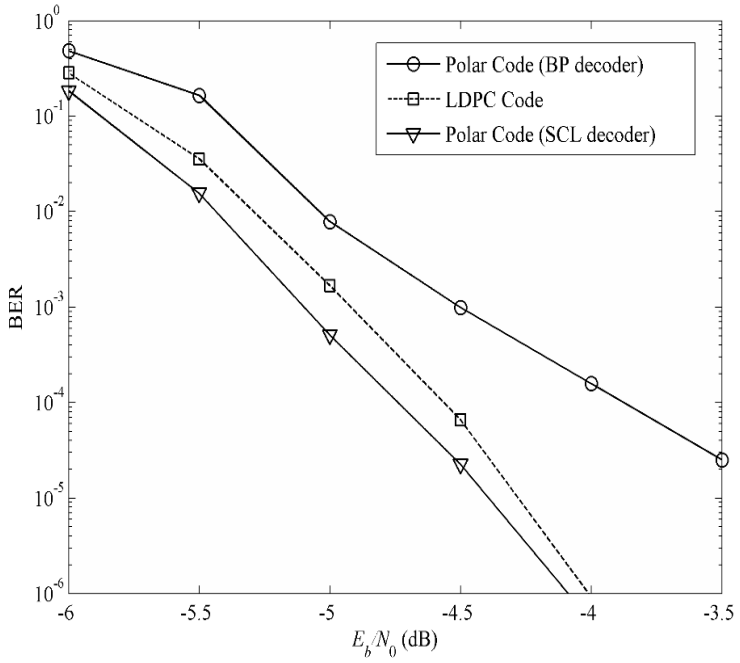


Fig. 4-11 The BER comparison for different coding schemes

4.5 Summary

In this chapter, we firstly introduced the basic principles and the mathematical model of the half-duplex single-relay channel. Then we studied in detail the polar coding technology under this model and proposed a novel cooperative partial message relay forward protocol. We elaborated on the coding method and briefly discussed the selection of partial messages in the CPMR protocol. By analyzing the asymptotic performance, we put forward that, while maintaining the low complexity of polar coding, the CPMR scheme can asymptotically achieve the capacity limit of the degraded half-duplex single-relay channel. In addition, the upper bound on the average block error probability under the SC decoding algorithm was also derived.

We also examined the practical polar coding scheme in the half-duplex relay system and its simulation performance. Based on "Turbo" theory, we designed a joint iterative soft parallel interference cancellation receiver structure based on polar codes using MAP detection at the receiver in order to recover the superposed information from the source node and the relay node in the MAC phase. Through the analysis of the relay channel capacity limit, we explained the key factors affecting the capacity limit of the channel model, based on which, we conducted simulation verification of the CPMR scheme of finite block lengths for the AWGN channel with BPSK modulation. The simulation results show that the proposed scheme can achieve a better performance than the conventional LDPC coding scheme under the SCLD decoding algorithm.

Chapter 5 Polar Coding Schemes for Multiple-Relay Transmission Systems

5.1 Introduction

From the research related to the previous chapter, we know that Meulen [13] introduced the classic three-node relay channel model, whose mathematical model is as shown in Fig. 5-1 [94]. In this model, node 1, which is a relay node, only serves to help transmit information from node 0 to node 2. A straightforward application of the model can be found in cooperative communications, where by using one node between the source node and the destination node can the distance be shortened by "one hop". Such applications involve studies on information payload, interference, power consumption and other issues.

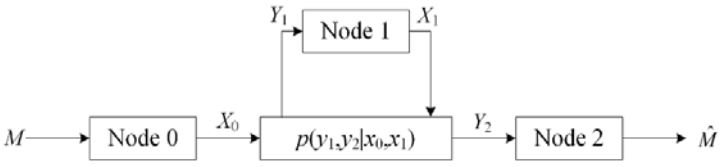


Fig. 5-1 The mathematical model of the general one-relay network

Fig. 5-1 demonstrates the simplest discrete memoryless single-relay network, where node 0, node 1 and node 2 respectively represent the source, relay and destination. This network model is denoted by $(\mathcal{X}_0 \times \mathcal{X}_1, p(y_1, y_2 \mid x_0, x_1), \mathcal{Y}_1 \times \mathcal{Y}_2)$, where $\mathcal{X}_0$ and $\mathcal{X}_1$ respectively represent the symbols transmitted by node 0 and node 1 while $\mathcal{Y}_1$ and $\mathcal{Y}_2$ denote the received symbols of node 0 and node 1. Specifically, the random variable X_0 is the channel input for source node 0, Y_2 the channel output for destination node 2 and Y_1 the output received by relay node 1. After processing the received Y_1, relay node 1 sends the input variable X_1. For an arbitrary time *t*, X_1 is a function expressed as follows:

$$x_1(t) = f_t(y_1(1), y_1(2), \ldots y_1(t-1)), \tag{5-1}$$

where $f_t(\cdot)$ is an arbitrary function and $(y_1(1), y_1(2), \ldots y_1(t-1))$ denotes the sequence received by the relay node before time *t*.

Up till now, research on the capacity of single-relay network still cannot exceed the maximum achievable rate derived and proved by Cover and Gamal [77]. The most important conclusion from their research regarding the DF protocol is that the following rate is achievable:

$$R_S = \sup_{p(x_0, x_1)} \min\{I(X_0; Y_1 \mid X_1), I(X_0, X_1; Y_2)\}. \tag{5-2}$$

In addition, if this relay network can be written in the following form

$$p(y_1, y_2 \mid x_0, x_1) = p(y_1 \mid x_0, x_1)\, p(y_2 \mid y_1, x_0, x_1) \tag{5-3}$$

$$= p(y_1 \mid x_0, x_1)\, p(y_2 \mid y_1, x_1), \tag{5-4}$$

then this relay network is called physically degraded. Equivalently, if the equalities in Equation (5-3) and Equation (5-4) hold, then $X_0 \to (X_1, Y_1) \to Y_2$ constitutes a Markov chain. At this point, the expression on the right-hand side of the equality in Equation (5-2) is exactly the capacity of a physically degraded single-relay network.

It should be noted that the rate in Equation (5-2) is achievable for any single-relay network. The prerequisite of being physically degraded only serves to prove that such a rate does not go beyond the boundary, and thus when the network has characteristics of physical degradation, it can achieve its system capacity. The rate can be put into practical application because it provides the achievable rate of the relay network, i.e., the lower bound on the network capacity. When the source-relay channel is better than the relay-destination channel (i.e., when the relay is physically closer to the source than to the destination), this lower bound is very compact and is arbitrarily approaching the network capacity. In order to achieve this rate, which is the achievable rate of the

DF relay protocol, the relay node needs a coding scheme for decoding the received source signal, and then delivering them to the destination node after re-encoding.

Based on cut-set theory of network information flow [91], Cover and Gamal also presented the upper bound on the system capacity of this single-relay network as follows [7]:

$$C_S \le \sup_{p(x_0,x_1)} \min(I(X_0, X_1; Y_2), I(X_0; Y_1, Y_2 \mid X_1)) . \tag{5-5}$$

If a single-relay network is one with orthogonal receiver components (SRN-ORCs), then its probability mass function describes the characteristics of the network as follows [138]:

$$p(y_1, y_2 \mid x_0, x_1) = p(y_{0,2}, y_1 \mid x_0)\, p(y_{1,2} \mid x_1) , \tag{5-6}$$

where $Y_2 = (Y_{0,2}, Y_{1,2})$. Fig. 5-2 describes this network system model, including the following vector of length N: $\boldsymbol{M}$ to be sent by the source includes information bits and frozen bits (the decoder is already known to the frozen bits) while $\hat{\boldsymbol{M}}$ is the corresponding estimation of $\boldsymbol{M}$. $\boldsymbol{X}_0$ is the symbol vector input from the source to the channel while $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_{0,2}$ are respectively the channel outputs of the relay node and the destination node. Similarly, $\boldsymbol{X}_1$ is the vector input by the relay node while $\boldsymbol{Y}_{1,2}$ is the value observed by the destination node. Define $W_{0,1}$, $W_{0,2}$ and $W_{1,2}$ respectively as the marginal probability mass functions of y_1, $y_{0,2}$ and $y_{1,2}$ in Equation (5-6).

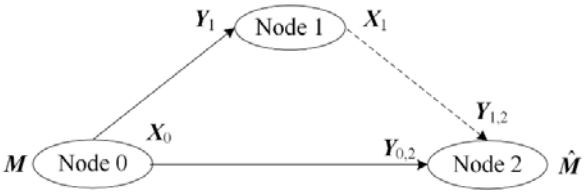


Fig. 5-2 The single-relay network with orthogonal receiver components

Before discussing about the capacity limit of this network model, we will first review the definition of degradation raised in the last chapter, which exactly explains

the statement that some channels are not as good as others.

Definition 5-1 (Stochastically Degradation): Let the channel $W_1 : \mathcal{X} \to \mathcal{Y}_1$ and the channel $W_2 : \mathcal{X} \to \mathcal{Y}_2$ be two B-DMCs. If W_1 is the stochastically degraded channel of W_2, then for each $y_1 \in \mathcal{Y}_1$ and $x \in \mathcal{X}$ there exists a B-DMC $W : \mathcal{Y}_2 \to \mathcal{Y}_1$, satisfying

$$W_1(y_1 \mid x) = \sum_{y_2} W_2(y_2 \mid x) W(y_1 \mid y_2) . \tag{5-7}$$

If the source-destination channel $W_{0,2}(y_{0,2} \mid x)$ in a single-relay network is stochastically degraded with respect to the source-relay channel $W_{0,1}(y_1 \mid x)$, then this network is called stochastically degraded. Similarly, if SRN-ORCs satisfies

$$p(y_{0,2}, y_2 \mid x) = W_{0,1}(y_1 \mid x)\, p(y_{0,2} \mid y_1) , \tag{5-8}$$

then this network is called physically degraded. From Equation (5-7) and Equation (5-8), we know that physical degradation implies random degradation.

In order to further explore the feasibility of applying the low complex CPMR scheme to multiple-relay systems, in this chapter, we carry out relevant research based on the primitive relay channel [138], which not only simplifies the system design but also makes it easier to explain the principles of the CPMR scheme. When determining the achievable rate during the application of the CPMR scheme based on polar codes in a primitive single-relay channel, the following additional assumptions are needed:

- $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ must satisfy binary uniform distribution;
- $W_{0,1}$, $W_{0,2}$ and $W_{1,2}$ must be symmetric channels;
- $W_{0,2}$ is the degraded channel of $W_{0,1}$.

Consider these requirements and then we have the following definitions.

Definition 5-2 (Upper Capacity Bound on SRN-ORCs):

$$C_S^O \le \max_{p(x_0)p(x_1)} \min\{I(X_0; Y_{0,2}) + I(X_1; Y_{1,2}), I(X_0; Y_1, Y_{0,2})\} . \tag{5-9}$$

Definition 5-3 (DF Achievable Rate of SRN-ORCs):

$$R_S^O = \min\{I(W_{0,2}) + I(W_{1,2}), I(W_{0,1})\}. \tag{5-10}$$

Based on the above two definitions, we consider the SRN-ORCs model, and for an arbitrary transmission rate $R < R_S^O$ there exists a polar code sequence of block length *N*, with the block error probability $\Pr(\hat{\boldsymbol{M}} \neq \boldsymbol{M})$ under the SC decoding algorithm to be upper bounded by $P_e \leq O(2^{-N^\beta})$, where $0 < \beta < 1/2$ [108].

In a practical wireless environment, there exists not just a single relay node in a wireless network, but rather probably multiple relay nodes assisting transmit information from the source node to the destination node. In fact, research on multiple-relay networks has already presented quite a few findings [91-96], but most of the work reached its conclusions based on non-constructive random coding schemes. These results indicate the existence of a coding sequence capable of achieving the capacity of a multiple-relay network but they do not address as to how to design a channel coding method with low encoding and decoding complexity. So far there are no relevant research findings on practical constructive channel coding methods capable of reaching the capacity of a multiple-relay network.

Since non-constructive random coding schemes have high complexity in multiple-relay networks, the main idea of this chapter is to extend our low-complexity CPMR scheme to multiple-relay networks with the aim of achieving the capacity of multiple-relay networks. Based on Razaghi's research on multiple-relay networks [96], this chapter first elaborates on the CPMR schemes for two-relay network systems, with the design concept being to describe and summarize the integral components of the CPMR scheme for multiple-relay networks. In this scheme, the message to be delivered by the relay node is the partial message from the previous relay node or the source node

(i.e., it corresponds to the partial information sets). It should be noted that the different relations between the messages to be delivered and their partial message sets will result in different CPMR schemes for multiple-relay networks. Therefore, in this chapter, we design CPMR schemes correspondingly for two types of degraded multiple-relay networks with orthogonal receiver components, propose the algorithm for computing the partial message set, analyze the constructive polar coding method, verify the asymptotic capacity achievability of the CPMR scheme and lastly derive the upper bound on the average block error probability.

5.2 Model of Multiple-relay Networks

Fig. 5-3 [94] presents the system model of a general multiple-relay network. The model describes a network with $K+2$ nodes, where the source node is marked as index 0, the destination node as index $K+1$ and the relay nodes sequentially as index 1 to *K*. The relay nodes themselves do not send any information, but simply help transmit messages from the source node to the destination node to address the uncertainty about source messages. We respectively define the random variable X_0 as the network input, Y_{K+1} as the ultimate output of the network, $X_1, X_2, \ldots, X_K$ as relay input and $Y_1, Y_2, \ldots, Y_K$ as the output of the corresponding relay node. This network model can be characterized by the joint probability mass function $p(y_1, y_2, \ldots, y_{K+1} \mid x_0, x_1, \ldots, x_K)$.

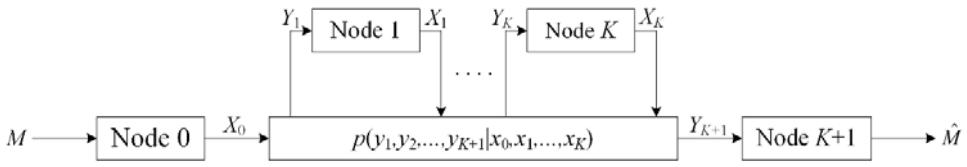


Fig. 5-3 The mathematical model of the general multiple relay network

The upper bound on the cut-set of this multiple-relay network model is [7]

$$C_M \le \max_{p(x_0,x_1,\ldots,x_K)} \min_{1\le k\le K+1} I(X_0,\ldots,X_{k-1};Y_k,\ldots,Y_{K+1} \mid X_k,\ldots,X_K). \quad (5\text{-}11)$$

In the case of a discrete memoryless multiple-relay network, if for an arbitrary $k=1,\ldots,K$, it satisfies:

$$p(y_{k+1},\ldots,y_{K+1} \mid y_k,x_0,\ldots,x_{k-1},x_k,\ldots,x_K) = p(y_{k+1},\ldots,y_{K+1} \mid y_k,x_k,\ldots,x_K), \quad (5\text{-}12)$$

then this multiple-relay network is called physically degraded [94]. Equivalently, Equation (5-12) shows that $(X_0,\ldots,X_{k-1}) \to (Y_k,X_k,\ldots,X_K) \to (Y_{k+1},\ldots,Y_{K+1})$ constitutes a Markov chain.

Based on the above multiple-relay network, Razaghi designed theoretically random coding methods and corresponding brand-new DF transmission protocols - PF protocol A and PF protocol B, for two special cases of this model (i.e., the serially-degraded multiple-relay network and the doubly-degraded multiple-relay network), and proved that this scheme can ensure reliable communication with a arbitrarily small error probability [96].

The channel capacity of a serially-degraded multiple-relay network is

$$C_M^A = \max_{p(x_0,\ldots,x_K)} \min_{1\le k\le K+1} I(X_0,\ldots,X_{k-1};Y_k \mid X_k,\ldots,X_K). \quad (5\text{-}13)$$

Razaghi proved that PF protocol A is capable of achieving this network capacity [96]. However, when the relay has a very short transmission distance and the message from the source node to the relay node is blocked due to a long distance, using PF protocol B can obtain a network capacity higher than C_M^A.

For a discrete memoryless doubly-degraded multiple-relay network model, if for an arbitrary $0 \le k \le K$, $X_k \to (Y_{k+1},X_{k+1}^K) \to Y_{k+2}^{K+1}$ and $X_0^{k-1} \to (Y_{K+1},X_k^K) \to Y_{k+1}^K$ constitute a Markov chain, then the capacity of the doubly-degraded multiple-relay network is

$$C_M^B = \max_{p(x_0,\ldots,x_K)} \min\{I(X_0;Y_1 \mid X_1^K),$$

$$I(X_0;Y_{K+1}\mid X_1^K)+I(X_1;Y_2\mid X_2^K),$$

$$I(X_0^1;Y_{K+1}\mid X_2^K)+I(X_2;Y_3\mid X_3^K),$$

$$\vdots$$

$$I(X_0^{K-2};Y_{K+1}\mid X_{K-1}^K)+I(X_{K-1};Y_K\mid X_K),$$

$$I(X_0^K;Y_{K+1})\}, \tag{5-14}$$

where $X_i^j=(X_i,X_{i+1},\ldots,X_j)$.

5.3 Model of Degraded MRN-ORCs

From the previous discussions, we know that there are some restrictions on directly applying polar codes to a multiple-relay network. Therefore, in this section, we focus on two types of degraded multiple-relay networks with orthogonal receiver components and directly deduce their capacities according to Equation (5-13) and Equation (5-14).

(1) Serially-Degraded Multiple-Relay Network with Orthogonal Receiver Components (SDMRN-ORCs).

Definition 5-4 (SDMRN-ORCs): Consider a serially-degraded multiple-relay network as shown in Fig. 5-4. If the channel output of this network can be further expressed as $Y_2=(Y_{0,2},Y_{1,2}),\ldots,Y_{K+1}=(Y_{0,K+1},Y_{1,K+1},\ldots,Y_{K,K+1})$, then we say that this network has orthogonal receiver components. Let $\{W_{k,l}\}$ be the $K+1$ B-DMC sets connecting the node *k* and the remaining $K+1-k$ nodes in the multiple-relay network, where $0\le k\le K$ and $k+1\le l\le K+1$. If $\{W_{k,K+1}\preceq W_{k,K}\preceq\cdots\preceq W_{k,1}\}$ is satisfied, then this network is called serially stochastically degraded.

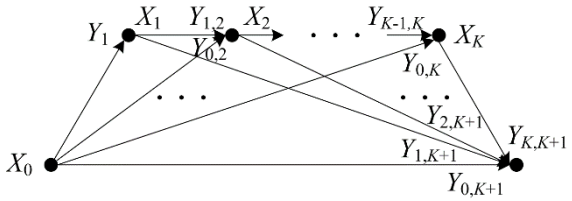


Fig. 5-4 The system model of SDMRN-ORCs

Lemma 5-1 (Transmission Properties of the Information Sets in SDMNR-ORCs): Consider a multiple-relay network model defined as above. If $\{W_{k,K+1} \preceq W_{k,K} \preceq \cdots \preceq W_{k,1}\}$ is satisfied, then for all the corresponding polarized bit channel indices *i*, there is $\{W_{k,K+1}^{(i)} \preceq W_{k,K}^{(i)} \preceq \cdots \preceq W_{k,1}^{(i)}\}$ and we have

$$\{Z(W_{k,K+1}^{(i)}) \geq Z(W_{k,K}^{(i)}) \geq \cdots \geq Z(W_{k,1}^{(i)})\}. \tag{5-15}$$

In addition, for $0 \leq k \leq K$, $k+1 \leq l \leq K+1$ and arbitrary $\delta > 0$, there exists information sets $\{\mathcal{A}_{k,l}\} \subseteq [N]$, satisfying

$$\{\mathcal{A}_{k,l}\} = \{i \in [N] : Z(W_{k,l}^{(i)}) \in [0, \delta)\}. \tag{5-16}$$

Then it has the following transmission properties:

$$\{\mathcal{A}_{k,K+1} \subseteq \mathcal{A}_{k,K} \subseteq \cdots \subseteq \mathcal{A}_{k,1}\}. \tag{5-17}$$

Proof: Let $0 < \beta < 1/2$ be an arbitrary constant and *N* be the polar code block length. Let the Bhattacharyya parameters of the polarized bit channel $W_{k,K+1}^{(i)}, W_{k,K}^{(i)}, \cdots, W_{k,1}^{(i)}$ be $Z(W_{k,K+1}^{(i)}), Z(W_{k,K}^{(i)}), \cdots, Z(W_{k,1}^{(i)})$ respectively. Let $\delta = \frac{1}{N} 2^{-N^{\beta}}$ (i.e., δ is arbitrarily small) and let $\mathcal{A}_{k,l+1}$ and $\mathcal{A}_{k,l}$ denote information sets:

$$\mathcal{A}_{k,l+1} = \{i \in [N] : Z(W_{k,l+1}^{(i)}) \leq \delta\}, \tag{5-18}$$

$$\mathcal{A}_{k,l} = \{i \in [N] : Z(W_{k,l}^{(i)}) \leq 1 - \delta^2\}. \tag{5-19}$$

For an arbitrarily small δ, there always exists $\delta \leq 1 - \delta^2$ and as the channel $W_{k,l+1}^{(i)}$ is the degraded channel of the channel $W_{k,l}^{(i)}$, $Z(W_{k,l+1}^{(i)}) \geq Z(W_{k,l}^{(i)})$ is satisfied. Therefore, for an arbitrary $i \in \mathcal{A}_{k,l+1}$, there exists the following relation:

$$Z(W_{k,l}^{(i)}) \le Z(W_{k,l+1}^{(i)}) \le \delta \le 1-\delta^2 . \tag{5-20}$$

This means that if $i \in \mathcal{A}_{k,l+1}$, then $i \in \mathcal{A}_{k,l}$, i.e., we can get $\mathcal{A}_{k,l+1} \subseteq \mathcal{A}_{k,l}$.

Therefore, all $0 \le k \le K$ and $k+1 \le l \le K+1$ satisfy Equation (5-15) and (5-17). □

As can be seen from the proof of Lemma 5-1, the transmission properties of information sets are a common phenomenon in degraded multiple-relay networks. When channel polarization occurs, all the "good" information bit indices of the degraded channel $W_{k,l}$ are also "good" information bit indices for the non-degraded channel $W_{k,l-1}, W_{k,l-2}, \ldots, W_{k,1}$, as shown by Equation (5-17). However, all the "good" information bit indices for the channel $W_{k,l-1}, W_{k,l-2}, \ldots, W_{k,1}$ are not necessarily "good" information bit indices for the degraded channel $W_{k,l}$.

Although the capacity of SDMNR-ORCs is still not known, we can easily deduce the achievable rate of SDMNR-ORCs from the right-hand side of the equality in Equation (5-13).

Corollary 5-1 (Achievable Rate of SDMNR-ORCs):

$$\begin{aligned} R_M^{OA} \le \max_{p(x_0)p(x_1)\cdots p(x_K)} \min\{ & I(X_0;Y_1), \\ & I(X_0;Y_{0,2}) + I(X_1;Y_{1,2}), \\ & \sum_{k=1}^{3} I(X_{k-1};Y_{k-1,3}), \\ & \vdots \\ & \sum_{k=1}^{K} I(X_{k-1};Y_{k-1,K}), \\ & \sum_{k=1}^{K+1} I(X_{k-1};Y_{k-1,K+1}) \} . \end{aligned} \tag{5-21}$$

If this network satisfies Equation (5-12), then the maximum achievable rate on

$p(x_0)p(x_1)\cdots p(x_K)$ is the capacity C_M^{OA}.

When there is only one relay node in this network, i.e., $K=1$, then Corollary 5-1 is simplified to be the capacity of the degraded SRN-ORCs, as shown by the right-hand side of the equality in Equation (5-10).

(2) Doubly-Degraded Multiple-Relay Network with Orthogonal Receiver Components (DDMRN-ORCs).

In this model, node *k* is completely blocked due to its long distance to node $k+2$ and the remaining relay nodes, so the source node can only communicate with relay node *k* through the relay nodes before relay node *k*.

Definition 5-5 (DDMRN-ORCs): Consider a doubly-degraded multiple-relay network as shown in Fig. 5-5. If the channel output of this network can be further expressed as $Y_{K+1}=(Y_{0,K+1},Y_{1,K+1},\ldots,Y_{K,K+1})$, then we say that this network has orthogonal receiver components. Let $\{W_{k,k+1},W_{k,K+1}\}$ be the *K* B-DMC sets connecting node *k* and node $k+1$ and connecting node *k* and $K+1$, where $0\le k\le K$. If $\{W_{k,K+1}\preceq W_{k,k+1}\}$, then the network is called doubly stochastically degraded.

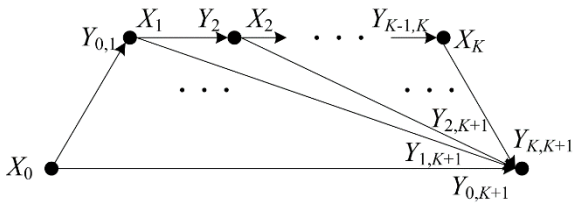


Fig. 5-5 The system model of DDMRN-ORCs

Lemma 5-2 (Transmission Properties of the Information Sets of DDMNR-ORCs): Consider a multiple-relay network defined as above. If $\{W_{k,K+1}\preceq W_{k,k+1}\}$ exists, then for all the corresponding polarized bit channel indices *i*, there is $\{W_{k,K+1}^{(i)}\preceq W_{k,k+1}^{(i)}\}$, so

$$\{Z(W_{k,K+1}^{(i)})\ge Z(W_{k,k+1}^{(i)})\}. \tag{5-22}$$

In addition, for $0 \le k \le K$, the subscript $l = \{k+1, K+1\}$ and an arbitrary $\delta > 0$, there exists information sets $\{\mathcal{A}_{k,l}\} \subseteq [N]$, satisfying

$$\{\mathcal{A}_{k,l}\} = \{i \in [N] : Z(W_{k,l}^{(i)}) \in [0, \delta)\}. \tag{5-23}$$

Then it has the following transmission properties:

$$\{\mathcal{A}_{k,K+1} \subseteq \mathcal{A}_{k,k+1}\}. \tag{5-24}$$

Proof: The proof can be deduced from the conclusion of Lemma 5-1. □

Similarly, based on Equation (5-14), we get the following corollary.

Corollary 5-2 (Achievable Rate of DDMNR-ORCs):

$$\begin{aligned} R_M^{OB} \le \max_{p(x_0)p(x_1)\cdots p(x_K)} \min\{ & I(X_0; Y_1), \\ & I(X_0; Y_{0,K+1}) + I(X_1; Y_2), \\ & I(X_0; Y_{0,K+1}) + I(X_1; Y_{1,K+1}) + I(X_2; Y_3), \\ & \vdots \\ & \sum_{k=1}^{K-1} I(X_{k-1}; Y_{k-1,K+1}) + I(X_{K-1}; Y_K), \\ & \sum_{k=1}^{K+1} I(X_{k-1}; Y_{k-1,K+1}) \}. \end{aligned} \tag{5-25}$$

In this network, if for $0 \le k \le K$, $X_k \to (Y_{k+1}, X_{k+1}^K) \to Y_{k+2}^{K+1}$ and $X_0^{k-1} \to (Y_{K+1}, X_k^K) \to Y_{k+1}^K$ constitute a Markov chain, then the maximum achievable rate on $p(x_0)p(x_1)\cdots p(x_K)$ is the capacity C_M^{OB}.

5.4 CPMR Scheme for Degraded TRN-ORCs

In this chapter, we propose a novel CPMR scheme employing the polar coding algorithm and prove that this method can achieve the network capacities C_M^{OA} and C_M^{OB}.

Different to the scheme in Ref. [92-96], our proposal no longer uses the block Markov superposition coding and the random "binning" method. Instead, we use a more practical polar coding method with low constructive complexity, which can be easily extended into multiple-relay networks. In order to make the capacity achievability easier to understand and also for a better comparison with Razaghi's coding method, firstly, we employ the serially-degraded two-relay network with orthogonal receiver components (SDTRN-ORCs) (where, $Y_2 = (Y_{0,2}, Y_{1,2}), Y_3 = (Y_{0,3}, Y_{1,3}, Y_{2,3})$) shown in Fig. 5-6 and the doubly-degraded two-relay network with orthogonal receiver components (DDTRN-ORCs) (where, , $Y_3 = (Y_{0,3}, Y_{1,3}, Y_{2,3})$) shown in Fig. 5-7 (i.e., $K = 2$. It should be noted that from the analysis of $K = 2$ hereinafter, we can directly get the coding scheme when $K = 1$) and prove the achievability of their capacities. This method of proof almost embodies all the basic ideas contributing to proving that the CPMR scheme can be used to achieve the capacity of this type of networks. Therefore, this method of proof can be directly extended to the case of $K \geq 2$.

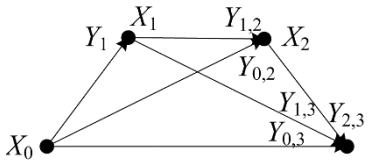


Fig. 5-6 The model of SDTRN-ORCs

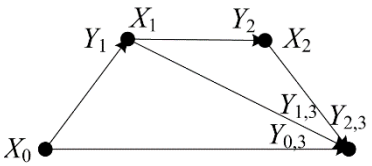


Fig. 5-7 The model of DDTRN-ORCs

From Corollary 5-1 and Corollary 5-2, we can easily get the network capacity of TRN-ORCs.

Corollary 5-3 (Capacity of the SDTNR-ORCs):

$$C_T^{OA} = \max_{p(x_0)p(x_1)p(x_2)} \min\{I(X_0; Y_1),$$

$$I(X_0;Y_{0,2})+I(X_1;Y_{1,2}),$$

$$I(X_0;Y_{0,3})+I(X_1;Y_{1,3})+I(X_2;Y_{2,3})\}. \quad (5\text{-}26)$$

Corollary 5-3 (Capacity of the DDTRN-ORCs):

$$C_T^{OB}=\max_{p(x_0)p(x_1)p(x_2)}\min\{I(X_0;Y_1),$$

$$I(X_0;Y_{0,3})+I(X_1;Y_2),$$

$$I(X_0;Y_{0,3})+I(X_1;Y_{1,3})+I(X_2;Y_{2,3})\}. \quad (5\text{-}27)$$

5.4.1 CPMR Transmission Strategy for Degraded TRN-ORCs

Different to the case of PF protocol A and PF protocol B in [96], for the two types of network models, i.e., SDTRN-ORCs and DDTRN-ORCs, in this chapter we respectively describe two CPMR protocols, namely the CPMR-A protocol and the CPMR-B protocol. Let $\boldsymbol{m}_0$ be the source message vector while $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ are respectively the message vectors forwarded by relay node 1 and relay node 2. The two protocols are different in that in the CPMR-A protocol, relay node 1 can decode the received source message $\boldsymbol{m}_0$ and then forward the partial message $\boldsymbol{m}_1$ of $\boldsymbol{m}_0$ to help relay node 2 decode the received source message $\boldsymbol{m}_0$; with the help of the received $\boldsymbol{m}_1$, relay node 2 can decode the source message $\boldsymbol{m}_0$, and then forward the partial message $\boldsymbol{m}_2$ of $\boldsymbol{m}_0$ to the destination node; with the help of the received partial messages $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$, the destination node can decode the received source message $\boldsymbol{m}_0$. In comparison, in the CPMR-B protocol, relay node 1 also decodes the received source message $\boldsymbol{m}_0$ and then forwards the partial message $\boldsymbol{m}_1$ of $\boldsymbol{m}_0$ to relay node 2; however, relay node 2 only decodes the received $\boldsymbol{m}_1$ instead of $\boldsymbol{m}_0$ and then forwards the partial message $\boldsymbol{m}_2$ of $\boldsymbol{m}_1$ to the destination node; with the help of $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$

(both the functions of $\boldsymbol{m}_0$), the destination node decodes the source message $\boldsymbol{m}_0$.

From Definition 5-4, we know that there are three communication channel sets in the SDTRN-ORCs model, $\{W_{0,1}, W_{0,2}, W_{0,3}\}$, $\{W_{1,2}, W_{1,3}\}$, and $\{W_{2,3}\}$, they satisfy the relations $W_{0,3} \preceq W_{0,2} \preceq W_{0,1}$ and $W_{1,3} \preceq W_{1,2}$. Let $\{\mathcal{A}_{0,1}, \mathcal{A}_{0,2}, \mathcal{A}_{0,3}\}$, $\{\mathcal{A}_{1,2}, \mathcal{A}_{1,3}\}$ and $\{\mathcal{A}_{2,3}\}$ be the corresponding three information sets (as shown in Lemma 5-1) and satisfy the relations $\mathcal{A}_{0,3} \subseteq \mathcal{A}_{0,2} \subseteq \mathcal{A}_{0,1}$ and $\mathcal{A}_{1,3} \subseteq \mathcal{A}_{1,2}$. From Definition 5-5, we know that although DDTRN-ORCs does not have the channel $W_{0,2}$ and its corresponding information set $\mathcal{A}_{0,2}$, it still satisfies the transmission properties in Lemma 5-2.

Assume that the source delivers the message $\boldsymbol{m}_0$ (whose information bit index belongs to the set $\mathcal{A}_{0,1}$) to relay node 1, relay node 2 and destination node 3, but because of $W_{0,3} \preceq W_{0,2} \preceq W_{0,1}$, relay node 2 and destination node 3 can only respectively obtain the partial messages corresponding to the subsets $\mathcal{A}_{0,2}$ and $\mathcal{A}_{0,3}$ of $\mathcal{A}_{0,1}$. In addition, relay node 1 delivers the partial message $\boldsymbol{m}_1$ (whose information bit index belongs to set $\mathcal{A}_{1,2}$) to relay node 2 and destination node 3, but because of $W_{1,3} \preceq W_{1,2}$, the destination node can only obtain the partial message corresponding to the subset $\mathcal{A}_{1,3}$ of $\mathcal{A}_{1,2}$. Lastly, relay node 2 delivers the message $\boldsymbol{m}_2$ (whose information bit index belongs to set $\mathcal{A}_{2,3}$) to destination node 3 to completely solve the uncertainty about the source message for destination node 3.

Based on the above analysis, the polarized channel set [N] corresponding to source node 0 can be divided into four independent subsets, as shown in Fig. 5-8 (where nodes 0 - 3 respectively represent the source node, the two relay nodes and the destination node):

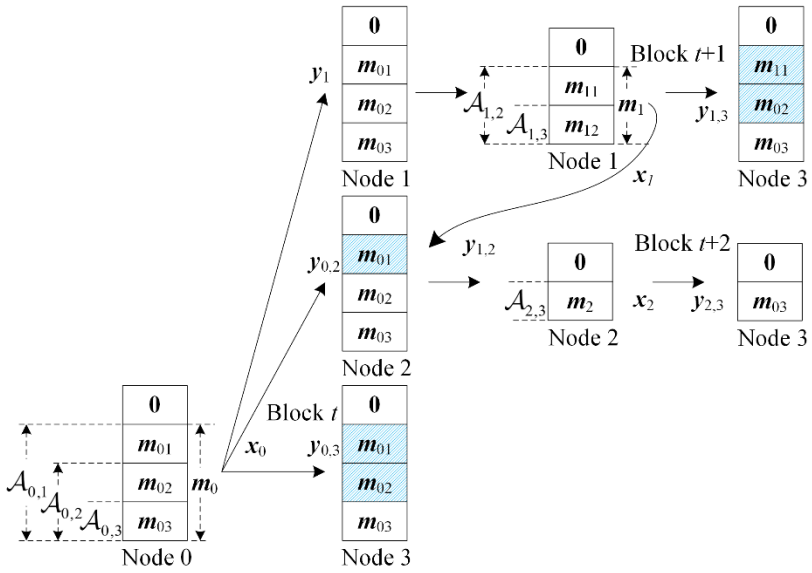


Fig. 5-8 The CPMR-A protocol for the SDTRN-ORCs model

- $\mathcal{A}_{0,1} \setminus \mathcal{A}_{0,2}$ represents the information set of the channel $W_{0,1}$ and the frozen set of the channels $W_{0,2}$ and $W_{0,3}$.
- $\mathcal{A}_{0,2} \setminus \mathcal{A}_{0,3}$ is the information set of the channels $W_{0,1}$ and $W_{0,2}$ and the frozen set of the channel $W_{0,3}$.
- $\mathcal{A}_{0,3}$ denotes information set of the channels $W_{0,1}$, $W_{0,2}$ and $W_{0,3}$;
- $\mathcal{F}_0 = [N] \setminus \mathcal{A}_{0,1}$ means the frozen set of the channels $W_{0,1}$, $W_{0,2}$ and $W_{0,3}$.

Accordingly, the set [N] of relay node 1 can be divided into three independent subsets:

- $\mathcal{A}_{1,2} \setminus \mathcal{A}_{1,3}$ is the information set of the channel $W_{1,2}$ and the frozen set of channel $W_{1,3}$.
- $\mathcal{A}_{1,3}$ denotes the information set of the channels $W_{1,2}$ and $W_{1,3}$;
- $\mathcal{F}_1 = [N] \setminus \mathcal{A}_{1,2}$ represents the frozen set of the channels $W_{1,2}$ and $W_{1,3}$.

Similarly, the set [N] of relay node 2 can be divided into two independent subsets:

- $\mathcal{A}_{2,3}$ represents the information set of the channel $W_{2,3}$;

- $\mathcal{F}_2 = [N] \setminus \mathcal{A}_{2,3}$ is the frozen set of the channel $W_{2,3}$.

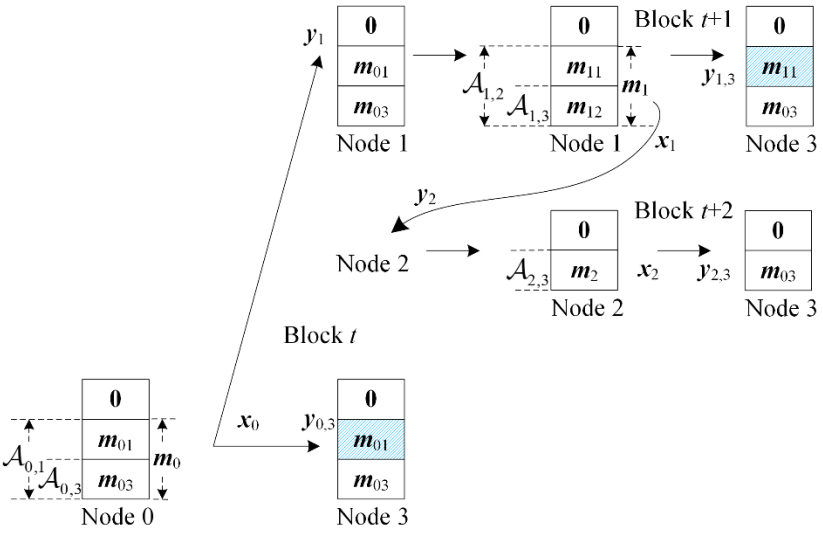


Fig. 5-9 The CPMR-B protocol for DDTRN-ORCs model

Similar to the CPMR-A protocol, the CPMR-B protocol divides the polarized channel set [*N*] corresponding to source node 0 in DDTRN-ORCs into just three independent subsets, as shown in Fig. 5-9:

- $\mathcal{A}_{0,1} \setminus \mathcal{A}_{0,3}$ represents the information set of the channel $W_{0,1}$ and the frozen set of the channel $W_{0,3}$.
- $\mathcal{A}_{0,3}$ is the information set of the channels $W_{0,1}$ and $W_{0,3}$,
- $\mathcal{F}_0 = [N] \setminus \mathcal{A}_{0,1}$ denotes the frozen set of the channels $W_{0,1}$ $W_{0,3}$.

The CPMR-B protocol uses the same partial message constructive method for relay node 1 and relay node 2 as the CPMR-A protocol.

Let *N* respectively represent the lengths of three time blocks as shown in Fig. 5-8. Because $W_{0,3} \preceq W_{0,2} \preceq W_{0,1}$, assume at time block *t*, the source node delivers the message bit vector $\boldsymbol{m}_0^{(t)}$, which is divided into three independent and disjoint sub message bit vectors $\boldsymbol{m}_{01}^{(t)} = \boldsymbol{m}_0^{(t)}(\mathcal{A}_{0,1} \setminus \mathcal{A}_{0,2})$ (the message bit vector which is in the set $\mathcal{A}_{0,1}$ but not in the set $\mathcal{A}_{0,2}$ of $\boldsymbol{m}_0^{(t)}$), $\boldsymbol{m}_{02}^{(t)} = \boldsymbol{m}_0^{(t)}(\mathcal{A}_{0,2} \setminus \mathcal{A}_{0,3})$ (the message bit vector

which is in the set $\mathcal{A}_{0,2}$ but not in the set $\mathcal{A}_{0,3}$ of $\boldsymbol{m}_0^{(t)}$) and $\boldsymbol{m}_{03}^{(t)} = \boldsymbol{m}_0^{(t)}(\mathcal{A}_{0,3})$ (the message bit vector which is in the set $\mathcal{A}_{0,3}$ of $\boldsymbol{m}_0^{(t)}$). At the same time, let $\boldsymbol{m}_1^{(t+1)}$ and $\boldsymbol{m}_2^{(t+2)}$ be the partial message vectors respectively forwarded by relay node 1 and relay node 2 at time blocks $t+1$ and $t+2$. In the CPMR-A protocol, the messages $\boldsymbol{m}_1^{(t+1)}$ and $\boldsymbol{m}_2^{(t+2)}$ delivered by the two relay nodes are respectively the previously received partial messages of the source message $\boldsymbol{m}_0^{(t)}$. $\boldsymbol{m}_1^{(t+1)}$ and $\boldsymbol{m}_2^{(t+2)}$ are computed from the following two equations:

$$\boldsymbol{m}_1^{(t+1)} = f_{\mathcal{B}_1}(\boldsymbol{m}_0^{(t)}), \tag{5-28}$$

$$\boldsymbol{m}_2^{(t+2)} = f_{\mathcal{B}_2}(\boldsymbol{m}_0^{(t)}). \tag{5-29}$$

In these equations, $f(\cdot)$ is the partial message function with the information sets of the source message and partial message as the parameters, where, $\mathcal{B}_1 = \mathcal{A}_{0,1} \setminus \mathcal{A}_{0,2}$ and $\mathcal{B}_2 = (\mathcal{A}_{1,2} \setminus \mathcal{A}_{1,3}) \bigcup (\mathcal{A}_{0,2} \setminus \mathcal{A}_{0,3})$.

In comparison, in the CPMR-B protocol as shown Fig. 5-9, the message $\boldsymbol{m}_0^{(t)}$ of source node 0 at time block t is only divided into two independent and disjoint sub message bit vectors $\boldsymbol{m}_{01}^{(t)} = \boldsymbol{m}_0^{(t)}(\mathcal{A}_{0,1} \setminus \mathcal{A}_{0,3})$ (the message bit vector which is in the set $\mathcal{A}_{0,1}$ but not in the set $\mathcal{A}_{0,3}$ of $\boldsymbol{m}_0^{(t)}$) and $\boldsymbol{m}_{03}^{(t)} = \boldsymbol{m}_0^{(t)}(\mathcal{A}_{0,3})$ (the message bit vector which is in the set $\mathcal{A}_{0,3}$ of $\boldsymbol{m}_0^{(t)}$). The messages $\boldsymbol{m}_1^{(t+1)}$ and $\boldsymbol{m}_2^{(t+2)}$ delivered by the two relay nodes are respectively the previously received partial message of the source message $\boldsymbol{m}_0^{(t)}$ and the partial message of $\boldsymbol{m}_1^{(t+1)}$ delivered by relay node 1. $\boldsymbol{m}_1^{(t+1)}$ and $\boldsymbol{m}_2^{(t+2)}$ are computed from the following two equations:

$$\boldsymbol{m}_1^{(t+1)} = f_{\mathcal{C}_1}(\boldsymbol{m}_0^{(t)}), \tag{5-30}$$

$$\boldsymbol{m}_2^{(t+2)} = f_{\mathcal{C}_2}(\boldsymbol{m}_1^{(t+1)}), \tag{5-31}$$

where $\mathcal{C}_1 = \mathcal{A}_{0,1} \setminus \mathcal{A}_{0,3}$ and $\mathcal{C}_2 = (\mathcal{A}_{1,2} \setminus \mathcal{A}_{1,3})$. In the CPMR-A protocol, the partial message sets $\mathcal{B}_1$ and $\mathcal{B}_2$ can be computed according to the algorithms in Table 5-1 to Table 5-3. While $\mathcal{C}_1$ and $\mathcal{C}_2$ in the CPMR-B protocol can also be computed based on such algorithms, the difference is that there do not exist the channel $W_{0,2}$ and the set $\mathcal{A}_{0,2}$ in the CPMR-B protocol. It should be noted that that the Bhattacharyya parameters of the channel $W_{1,k}, (k = \{2,3\})$ in relay node 1 can also be computed according to the algorithm in Table 5-2.

Table 5-1 The algorithm of computing the erasure probability for a given channel model

Input: A specific B-DMC channel model
Output: The "erasure probability" ε for the given channel model
If the channel is BEC **then**
  define $\varepsilon =$ "erasure probability"
elseif the channel is BSC **then**
  define $\varepsilon = 1 - H(p)$ //p is the transition probability for the BSC channel and $H(.)$ is the information entropy.
elseif the channel is BI-AWGN **then**
  define $\varepsilon = 1 - I(W)$ //$I(.)$ is the mutual information for the channel W
end if
return ε

Table 5-2 The algorithm of computing Bhattacharyya parameters for the source node

Input: The "erasure probability" $\varepsilon_{0,j}, (j = \{1,2,3\})$ for the B-DMC $W_{0,j}, (j = \{1,2,3\})$
Output: Bhattacharyya parameters.
for j = 1,2,3 **do**
  $Z(W_{0,j}) = \varepsilon_{0,j}$
  for 1 to N **do** //compute Bhattacharyya parameters
    $Z_{0,j}^{(1)} = 2Z(W_{0,j}^{(i)}) - Z(W_{0,j}^{(i)})^2$
    $Z_{0,j}^{(2)} = Z(W_{0,j}^{(i)})^2$
    $\{Z(W_{0,j}^{(i)})\} = [Z_{0,j}^{(1)}; Z_{0,j}^{(2)}]$
  end for
end for
return $\{Z(W_{0,j}^{(i)})\}$

Table 5-3 The algorithm of computing the partial message sets for Protocol CPMR-A

Input: $\{Z(W_{0,j}^{(i)})\}$ and $\{Z(W_{1,k}^{(i)})\}$

Output: Partial message sets

1 **for** j = 1,2,3 **do**

2 $[value_{0,j}, index_{0,j}] = sort(\{Z(W_{0,j}^{(i)})\}, \text{'descending'})$ // sort function sorts $\{Z(W_{0,j}^{(i)})\}$ in descending order and returns the values as $value_{0,j}$ and also returns a corresponding index matrix $index_{0,j}$

3 $K_{0,j} = NI(W_{0,j})$

4 $\mathcal{A}_{0,j} = index_{0,j}(N - K_{0,j} + 1 : N)$

5 **end for**

6 **for** k = 2,3 **do**

7 $[value_{1,k}, index_{1,k}] = sort(\{Z(W_{1,k}^{(i)})\}, \text{'descending'})$

8 $K_{1,k} = NI(W_{1,k})$

9 $\mathcal{A}_{1,k} = index_{1,k}(N - K_{1,k} + 1 : N)$

10 **end for**

11 $\mathcal{B}_1 = \mathcal{A}_{0,1} \setminus \mathcal{A}_{0,2} = setdiff(\mathcal{A}_{0,1}, \mathcal{A}_{0,2})$ //the setdiff function returns the difference between two sets

12 $\mathcal{B}_2 = (\mathcal{A}_{1,2} \setminus \mathcal{A}_{1,3}) \cup (\mathcal{A}_{0,2} \setminus \mathcal{A}_{0,3}) = setdiff(\mathcal{A}_{1,2}, \mathcal{A}_{1,3}) \cup setdiff(\mathcal{A}_{0,2}, \mathcal{A}_{0,3})$

13 **return** $\mathcal{B}_1$ and $\mathcal{B}_2$

5.4.2 Polar Encoding and Decoding Process for the Degraded TRN-ORCs

Table 5-4 The encoding and decoding process for the SDTRN-ORCs using CPMR-A protocol

Time	node 0	node 1		node 2		node 3
block	Send	Receive	Send	Receive	Send	Receive
1	$\boldsymbol{x}_0^{(1)}$	$\boldsymbol{y}_{0,1}^{(1)}$	$\emptyset$	$\boldsymbol{y}_{0,2}^{(1)}$	$\emptyset$	$\boldsymbol{y}_{0,3}^{(1)}$
2	$\boldsymbol{x}_0^{(2)}$	$\boldsymbol{y}_{0,1}^{(2)}$	$\boldsymbol{x}_1^{(1)}$	$\boldsymbol{y}_{1,2}^{(1)}, \boldsymbol{y}_{0,2}^{(2)}$	$\emptyset$	$\boldsymbol{y}_{1,3}^{(1)}, \boldsymbol{y}_{0,3}^{(2)}$
3	$\boldsymbol{x}_0^{(3)}$	$\boldsymbol{y}_{0,1}^{(3)}$	$\boldsymbol{x}_1^{(2)}$	$\boldsymbol{y}_{1,2}^{(2)}, \boldsymbol{y}_{0,2}^{(3)}$	$\boldsymbol{x}_2^{(1)}$	$\boldsymbol{y}_{2,3}^{(1)}, \boldsymbol{y}_{1,3}^{(2)}, \boldsymbol{y}_{0,3}^{(3)}$
				⋮		
t	$\boldsymbol{x}_0^{(t)}$	$\boldsymbol{y}_{0,1}^{(t)}$	$\boldsymbol{x}_1^{(t-1)}$	$\boldsymbol{y}_{1,2}^{(t-1)}, \boldsymbol{y}_{0,2}^{(t)}$	$\boldsymbol{x}_2^{(t-2)}$	$\boldsymbol{y}_{2,3}^{(t-2)}, \boldsymbol{y}_{1,3}^{(t-1)}, \boldsymbol{y}_{0,3}^{(t)}$
t+1	$\boldsymbol{x}_0^{(t+1)}$	$\boldsymbol{y}_{0,1}^{(t+1)}$	$\boldsymbol{x}_1^{(t)}$	$\boldsymbol{y}_{1,2}^{(t)}, \boldsymbol{y}_{0,2}^{(t+1)}$	$\boldsymbol{x}_2^{(t-1)}$	$\boldsymbol{y}_{2,3}^{(t-1)}, \boldsymbol{y}_{1,3}^{(t)}, \boldsymbol{y}_{0,3}^{(t+1)}$
t+2	$\boldsymbol{x}_0^{(t+2)}$	$\boldsymbol{y}_{0,1}^{(t+2)}$	$\boldsymbol{x}_1^{(t+1)}$	$\boldsymbol{y}_{1,2}^{(t+1)}, \boldsymbol{y}_{0,2}^{(t+2)}$	$\boldsymbol{x}_2^{(t)}$	$\boldsymbol{y}_{2,3}^{(t)}, \boldsymbol{y}_{1,3}^{(t+1)}, \boldsymbol{y}_{0,3}^{(t+2)}$
				⋮		

The encoding and decoding process in the CPMR-A protocol for the SDTRN-ORCs model is executed sequentially, following the time block order (see Table 5-4; in the

encoding and decoding using the CPMR-B protocol for the DDTRN-ORCs model, relay node 2 does not receive the source message delivered by the source node):

(1) Time block t

The source encoder firstly encodes the message $\boldsymbol{m}_0$ into a polar codeword $\boldsymbol{x}_0$ and then delivers it to relay node 1. As the frozen set $\mathcal{F}_0$ is already known to the relay nodes, relay node 1 can execute the SC decoding algorithm and thus reliably reconstruct the source message, when the transmission rate of the source message is

$$\lim_{N\to\infty} R_0 = \lim_{N\to\infty} \frac{|\mathcal{A}_{0,1}|}{N} = I(X_0;Y_1). \tag{5-32}$$

On the other hand, although in the CPMR-A protocol relay node 2 and destination node 3 receive the codeword $\boldsymbol{x}_0$ at the same time, relay node 2 only receives $NI(X_0;Y_{0,2})$ bits of information of the sub messages $\boldsymbol{m}_{02}$ and $\boldsymbol{m}_{03}$ in $\boldsymbol{x}_0$ but still lacks $N(I(X_0;Y_1)-I(X_0;Y_{0,2}))$ bits of information of the sub message $\boldsymbol{m}_{01}$ in $\boldsymbol{x}_0$ in order to fully recover the source message. Meanwhile, destination node 3 only receives $NI(X_0;Y_{0,3})$ bits of information of the sub message $\boldsymbol{m}_{03}$ in $\boldsymbol{x}_0$ but lacks $N(I(X_0;Y_1)-I(X_0;Y_{0,3}))$ bits of information of the sub messages $\boldsymbol{m}_{01}$ and $\boldsymbol{m}_{02}$ in order to fully recover the source message. Therefore, relay node 2 and the destination node only store the received $\boldsymbol{x}_0$ and do not conduct further decoding until they have received sufficient helpful partial messages. However, in the CPMR-B protocol, relay node 2 does not receive the source message but only receives the partial message from relay node 1.

(2) Time block $t+1$

The codeword to be delivered by relay node 1 is determined by the codeword $\boldsymbol{x}_0$ delivered by the source, so relay node 1 can cooperate with the source node in

transmitting the source message. Because in the CPMR-A protocol $\mathcal{A}_{0,3} \subseteq \mathcal{A}_{0,2} \subseteq \mathcal{A}_{0,1}$, relay node 1 in each time block knows exactly what partial message is required by relay node 2 to reliably recover the source message. Similar to the treatment of the source node, the partial message $\boldsymbol{m}_1$ can also be divided into two independent and disjoint sub vectors: $\boldsymbol{m}_{11} = \boldsymbol{m}_1(\mathcal{A}_{1,2} \setminus \mathcal{A}_{1,3})$ (the message bit vector which is in the set $\mathcal{A}_{1,2}$ but not in the set $\mathcal{A}_{1,3}$ of $\boldsymbol{m}_1$) and $\boldsymbol{m}_{12} = \boldsymbol{m}_1(\mathcal{A}_{1,3})$ (the message bit vector which is in the set $\mathcal{A}_{1,3}$ of $\boldsymbol{m}_1$). When the decoded source message $\boldsymbol{m}_0$ is given (assume the previous $\boldsymbol{m}_0$ decoded is correct), relay node 1 can generate a polar codeword $\boldsymbol{x}_1$ of block length *N*. Therefore, the partial message $\boldsymbol{m}_1$ of $\boldsymbol{m}_0^{(t)}$ decoded by relay node 1 (in fact, if the decoding by relay node 1 is correct, then $\boldsymbol{m}_1^{(t+1)} = \boldsymbol{m}_{01}^{(t)}$) is delivered to relay node 2 and destination node 3 to resolve the two nodes' uncertainty about the source message $\boldsymbol{m}_0^{(t)}$, with the transmission rate being

$$R_1 = \frac{|\mathcal{A}_{0,1}| - |\mathcal{A}_{0,2}|}{N}. \tag{5-33}$$

With the help of the received $\boldsymbol{m}_1^{(t+1)} = f_{\mathcal{B}_1}(\boldsymbol{m}_0^{(t)})$, relay node 2 can decode the source message $\boldsymbol{m}_0^{(t)}$. However, although the destination node has respectively obtained $NI(X_0;Y_{0,3})$ bits of information and $NI(X_1;Y_{1,3})$ bits of information during time block *t* and $t+1$, it still lacks $N(I(X_0;Y_1) - I(X_0;Y_{0,3}) - I(X_1;Y_{1,3}))$ bits of information for reconstructing the source message. In the CPMR-B protocol, because $\mathcal{A}_{0,3} \subseteq \mathcal{A}_{0,1}$ and $\mathcal{A}_{1,3} \subseteq \mathcal{A}_{1,2}$, relay node 1 knows what partial message is required by the destination node for reliable decoding. Therefore, relay node 1 delivers the partial message $\boldsymbol{m}_1$ to relay node 2 and destination node 3 to just resolve the destination node's uncertainty about the source message. With the help of $\boldsymbol{m}_1^{(t+1)} = f_{\mathcal{C}_1}(\boldsymbol{m}_0^{(t)})$, relay node 2 can also

correctly decode the message $\boldsymbol{m}_1$.

(3) Time block $t+2$

Relay node 2 encodes the message $\boldsymbol{m}_2$ into the codeword $\boldsymbol{x}_2$ of block length N, which is transmitted to the destination node at the rate

$$R_2 = \frac{|\mathcal{A}_{1,2}| - |\mathcal{A}_{1,3}| + |\mathcal{A}_{0,2}| - |\mathcal{A}_{0,3}|}{N} \tag{5-34}$$

to finally resolve the uncertainty about the source message $\boldsymbol{m}_0^{(t)}$. In fact, when both relay node 1 and relay node 2 correctly decode, $\boldsymbol{m}_2^{(t+2)} = (\boldsymbol{m}_{11}^{(t+1)}, \boldsymbol{m}_{02}^{(t)})$. Lastly, the destination node uses the received helpful messages $\boldsymbol{m}_1^{(t+1)}$ and $\boldsymbol{m}_2^{(t+2)}$ as the a priori information so as to reconstruct the source message $\boldsymbol{m}_0^{(t)}$. At this point, the codeword $\boldsymbol{x}_0$ is considered as the codeword $\boldsymbol{x}'_0$ delivered from the source node to the destination node during time block t, so the rate accordingly decreases to

$$\lim_{N\to\infty} R'_0 = \lim_{N\to\infty} \frac{|\mathcal{A}_{0,3}|}{N} = I(X_0; Y_{0,3}). \tag{5-35}$$

In the CPMR-B protocol, relay node 2 sends the partial message $\boldsymbol{m}_2^{(t+2)}$ (in fact, when both relay node 1 and relay node 2 can correctly decode, $\boldsymbol{m}_2^{(t+2)} = \boldsymbol{m}_{01}^{(t)} \setminus \boldsymbol{m}_{12}^{(t+1)}$, which represents the bits in $\boldsymbol{m}_{01}^{(t)}$ but not in $\boldsymbol{m}_{12}^{(t+1)}$) to the destination node to finally resolve the uncertainty about the source message $\boldsymbol{m}_0^{(t)}$. Lastly, the destination node uses the received helpful messages $\boldsymbol{m}_1^{(t+1)} = f_{\mathcal{C}_1}(\boldsymbol{m}_0^{(t)})$ and $\boldsymbol{m}_2^{(t+2)} = f_{\mathcal{C}_2}(\boldsymbol{m}_1^{(t+1)})$ as the a priori information so as to reconstruct the source message $\boldsymbol{m}_0^{(t)}$.

From the above discussion, we know that, when $N\to\infty$, both the CPMR-A protocol and the CPMR-B protocol based on polar codes can asymptotically achieve the capacities C_T^{OA} and C_T^{OB}. Therefore, the upper bound on the average block error probability for the CPMR scheme can be determined according to the following

theorem.

Theorem 5-1 (Capacity Achievability of Polar Codes for the SDTRN-ORCs and SDTRN-ORCs): Consider the SDTRN-ORCs and DDTRN-ORCs. Any transmission rate $0 < R < C_T^{OA}$ or $0 < R < C_T^{OB}$ is said to be achievable if there exists a sequence of polar codes of block length *N*, such that the block error probability $P_e = \Pr(\hat{\boldsymbol{m}}_0 \neq \boldsymbol{m}_0)$ at the destination under the SC decoding algorithm is bounded by $P_e \leq O(2^{-N^\beta})$ for any $0 < \beta < 1/2$ and sufficiently large *N*.

5.4.3 Asymptotic Performance of the Block Error Probability

This subsection proves Theorem 5-1. Let $\mathcal{E}_0$ represent the error event which occurs when destination node 3 reconstructs the source message $\boldsymbol{m}_0^{(t)}$. Let $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{E}_3$ respectively be the event $\{\hat{\boldsymbol{m}}_0 \neq \boldsymbol{m}_0\}$ occurring at relay node 1, the event $\{\hat{\boldsymbol{m}}_1 \neq \boldsymbol{m}_1\}$ occurring when relay node 2 recovers the partial message $\boldsymbol{m}_1$ delivered by relay node 1 and the event $\{\hat{\boldsymbol{m}}_2 \neq \boldsymbol{m}_2\}$ occurring when destination node 3 recovers the partial message $\boldsymbol{m}_2$ delivered by relay node 2. At the same time, define $\mathcal{E}_1^c$, $\mathcal{E}_2^c$ and $\mathcal{E}_3^c$ as the complementary events of the above events. According to the formula of total probability, the block error probability of the CPMR scheme can be computed as

$$\begin{aligned} P_e &= \Pr(\mathcal{E}_0) \\ &= \Pr(\mathcal{E}_0 \bigcap (\mathcal{E}_1 \bigcup \mathcal{E}_1^c)) \\ &\leq \Pr(\mathcal{E}_0 \bigcap \mathcal{E}_1) + \Pr(\mathcal{E}_0 \bigcap \mathcal{E}_1^c) \\ &= \Pr(\mathcal{E}_1)\Pr(\mathcal{E}_0 \mid \mathcal{E}_1) + \Pr(\mathcal{E}_1^c)\Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c) \\ &\leq \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c). \end{aligned} \tag{5-36}$$

If the polar code sequence is transmitted from source node 0 to relay node 1 at rate

$R_0 < I(X_0;Y_1)$, then based on the properties of polar codes, we can easily get

$$\Pr(\mathcal{E}_1) \le O(2^{-N^\beta}). \tag{5-37}$$

The second term in Equation (5-36) indicates the error probability for the destination node to reconstruct the source message $\boldsymbol{m}_0$ when relay node 1 decodes correctly. Therefore, this term can be further written into

$$\begin{aligned}\Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c) &\le \Pr(\mathcal{E}_2 \mid \mathcal{E}_1^c)\Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2) + \Pr(\mathcal{E}_2^c \mid \mathcal{E}_1^c)\Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c) \\ &\le \Pr(\mathcal{E}_2 \mid \mathcal{E}_1^c) + \Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c).\end{aligned} \tag{5-38}$$

The first term in Equation (5-38) indicates the error probability for relay node 2 to reconstruct the partial message $\boldsymbol{m}_1$ when relay node 1 decodes correctly. Therefore, when relay node 1 delivers the partial message to relay node 2 at a coding rate smaller than the channel capacity, i.e. $R_1 < I(X_1;Y_{1,2})$, then its error probability is upper bounded by

$$\Pr(\mathcal{E}_2 \mid \mathcal{E}_1^c) \le O(2^{-N^\beta}). \tag{5-39}$$

Similarly, the second term in Equation (5-38) can be further written into

$$\begin{aligned}\Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c) &\le \Pr(\mathcal{E}_3 \mid \mathcal{E}_1^c, \mathcal{E}_2^c)\Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \mathcal{E}_3) + \Pr(\mathcal{E}_3^c \mid \mathcal{E}_1^c, \mathcal{E}_2^c)\Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \mathcal{E}_3^c) \\ &\le \Pr(\mathcal{E}_3 \mid \mathcal{E}_1^c, \mathcal{E}_2^c) + \Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \mathcal{E}_3^c).\end{aligned} \tag{5-40}$$

For any $R_2 < I(X_2;Y_{2,3})$, it satisfies

$$\Pr(\mathcal{E}_3 \mid \mathcal{E}_1^c, \mathcal{E}_2^c) \le O(2^{-N^\beta}). \tag{5-41}$$

The second term in Equation (5-40) indicates the error probability for destination node 3 to reconstruct the source message $\boldsymbol{m}_0$ when destination node 3 receives two correct helpful partial messages $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ from the two relays. Therefore, for any $R'_0 < I(X_0;Y_3)$, its error probability is also upper bounded by

$$\Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \mathcal{E}_3^c) \le O(2^{-N^\beta}). \tag{5-42}$$

Therefore, based on Equation (5-37) and Equation (5-39) to (5-42), we know that when $N\to\infty$, the probability of the error event $\mathcal{E}_0$ is upper bounded by $O(2^{-N^\beta})$. □

5.5 CPMR Scheme for Degraded MRN-ORCs

By letting the message delivered by the relay node be the partial message corresponding to the information set of the partial message from the other relay nodes or the source node, the CPMR scheme discussed in the previous section can be extended to multiple-relay networks. As the different relations between the message forwarded by the relay node and the messages forwarded by its neighboring relay nodes will result in different CPMR protocols in a multiple-relay network, we will next deduce and prove that the CPMR scheme can also achieve the capacity of the SDMNR-ORCs model and that of the DDMNR-ORCs model.

5.5.1 CPMR Transmission Strategy for the Degraded MRN-ORCs

From Definition 5-4, it is known that for $0 \le k \le K$ and $k+1 \le l \le K+1$, we can let $\{W_{k,l}\}$ be the $K+1$ given channel sets in the SDMRN-ORCs model. Let $\{\mathcal{A}_{k,l}\} \subseteq [N]$ be the corresponding information set and satisfy the transmission properties of Lemma 5-1. From Definition 5-5, we know that for all $0 \le k \le K$, we can let $\{W_{k,k+1}, W_{k,K+1}\}$ respectively be the K given channel sets in the DDMRN-ORCs model. Let $\{\mathcal{A}_{k,K+1}, \mathcal{A}_{k,k+1}\}$ be the corresponding information set satisfying the transmission properties of Lemma 5-2.

Assume that in the CPMR-A protocol, source node 0 delivers the message $\boldsymbol{m}_0$ (whose information bit index belongs to the set $\mathcal{A}_{0,1}$) to the remaining $K+1$ nodes

and because $W_{0,K+1} \preceq W_{0,K} \preceq \cdots \preceq W_{0,1}$, these nodes can only respectively obtain the partial source information corresponding to the subset $\{\mathcal{A}_{0,l}\}$ of $\mathcal{A}_{0,1}$. In addition, relay node *k* forwards the partial message $\boldsymbol{m}_k$ (whose information bit index belongs to the set $\mathcal{A}_{k,k+1}$) to relay node $k+1$ and the remaining $K+1-(k+1)$ degraded channels, but because $W_{k,K+1} \preceq W_{k,K} \preceq \cdots \preceq W_{k,k+1}$, the remaining nodes can only respectively obtain the partial source messages corresponding to the subset $\{\mathcal{A}_{k,l}\}_{l=k+2}^{K+1}$ of $\mathcal{A}_{k,k+1}$. Lastly, relay node *K* delivers the message $\boldsymbol{m}_K$ (whose information bit index belongs to the set $\mathcal{A}_{K,K+1}$) to the destination node to fully resolve the destination node's uncertainty about the source message. Therefore, the polarized channel set [*N*] corresponding to source node 0 can be divided into $K+2$ independent and disjoint subsets:

- $\mathcal{A}_{0,l} \setminus \mathcal{A}_{0,l+1}$ represents the information set of the channel $W_{0,1},\ldots,W_{0,l}$ and the frozen set of the channel $W_{0,l+1},\ldots,W_{0,K+1}$;
- $\mathcal{A}_{0,K+1}$ is the information set of the channel $W_{0,1},\ldots,W_{0,K+1}$;
- $\mathcal{F}_0 = [N] \setminus \mathcal{A}_{0,1}$ denotes the frozen set of the channel $W_{0,1},\ldots,W_{0,K+1}$.

Accordingly, the set [*N*] of relay node *k* can be divided into three independent subsets:

- $\mathcal{A}_{k,l} \setminus \mathcal{A}_{k,l+1}$ is the information set of the channel $W_{k,l}$ and the frozen set of the channel $W_{k,l+1}$;
- $\mathcal{A}_{k,K+1}$ represents the information set of the channels $W_{k,l}$ and $W_{k,K+1}$;
- $\mathcal{F}_k = [N] \setminus \mathcal{A}_{k,l}$ means the frozen set of the channels $W_{k,l}$ and $W_{k,K+1}$.

Similarly, the set [*N*] of relay node *K* can be divided into two independent subsets:

- $\mathcal{A}_{K,K+1}$ represents the information set of the channel $W_{K,K+1}$;
- $\mathcal{F}_K = [N] \setminus \mathcal{A}_{K,K+1}$ denotes the frozen set of the channel $W_{K,K+1}$.

Similar to the discussion on the CPMR-A protocol, we also define the information sets and frozen sets for the CPMR-B protocol. The difference lies in that the relay nodes in the CPMR-B protocol have a very short transmission distance and the transmission is blocked due to the long channel distance between the source node and the relay nodes. Therefore, we divide the set [*N*] of the polarized bit channel corresponding to the source node in the DDMRN-ORCs model into just three independent and disjoint subsets:

- $\mathcal{A}_{0,1} \setminus \mathcal{A}_{0,K+1}$ represents the information set of the channel $W_{0,1}$ and the frozen set of the channel $W_{0,K+1}$;
- $\mathcal{A}_{0,K+1}$ is the information set of the channels $W_{0,1}$ and $W_{0,K+1}$;
- $\mathcal{F}_0 = [N] \setminus \mathcal{A}_{0,1}$ denotes the frozen set of the channels $W_{0,1}$ and $W_{0,K+1}$.

The set [*N*] of relay node *k* can be divided into three independent and disjoint subsets:

- $\mathcal{A}_{k,k+1} \setminus \mathcal{A}_{k,K+1}$ represents the information set of the channel $W_{k,k+1}$ and the frozen set of the channel $W_{k,K+1}$;
- $\mathcal{A}_{k,K+1}$ is the information set of the channels $W_{k,k+1}$ and $W_{k,K+1}$;
- $\mathcal{F}_k = [N] \setminus \mathcal{A}_{k,k+1}$ denotes the frozen set of the channels $W_{k,k+1}$ and $W_{k,K+1}$.

Let *N* respectively represent the length of each time block. Because $W_{0,K+1} \preceq W_{0,K} \preceq \cdots \preceq W_{0,1}$, assume that during time block *t*, source node 0 wants to send the information bit vector $\boldsymbol{m}_0^{(t)}$, which is divided into $K+1$ independent and disjoint sub information bit vectors $\boldsymbol{m}_{01}^{(t)},\ldots,\boldsymbol{m}_{0l}^{(t)},\ldots,\boldsymbol{m}_{0K+1}^{(t)}$, where $\boldsymbol{m}_{0l}^{(t)} = \boldsymbol{m}_0^{(t)}(\mathcal{A}_{0,l} \setminus \mathcal{A}_{0,l+1})$ (the information bit vector which is in the set $\mathcal{A}_{0,l}$ but not in the set $\mathcal{A}_{0,l+1}$ of $\boldsymbol{m}_0^{(t)}$). Also, let $\boldsymbol{m}_k^{(t+l)}$ be the partial message delivered by relay node *k* during time block $t+l$. In the CPMR-A protocol, the message $\boldsymbol{m}_k^{(t+l)}$ delivered by *K* relay nodes is the

corresponding partial message of the previously received source message $\boldsymbol{m}_0^{(t)}$. $\boldsymbol{m}_k^{(t+l)}$ is computed from the following equation:

$$\boldsymbol{m}_k^{(t+l)} = f_{\mathcal{B}_k}(\boldsymbol{m}_0^{(t)}), \tag{5-43}$$

where $f(\cdot)$ is the partial message function with the information sets of the source message $\boldsymbol{m}_0^{(t)}$ and the partial message as the parameters, where,

$$\begin{aligned}
&\mathcal{B}_1 = \mathcal{A}_{0,1} \setminus \mathcal{A}_{0,2}, \\
&\mathcal{B}_2 = (\mathcal{A}_{1,2} \setminus \mathcal{A}_{1,3}) \cup (\mathcal{A}_{1,2} \setminus \mathcal{A}_{1,3}), \\
&\quad \vdots \\
&\mathcal{B}_k = \bigcup_{k=1}^{K-2} (\mathcal{A}_{k,k+1} \setminus \mathcal{A}_{k,K}) \cup (\mathcal{A}_{0,k} \setminus \mathcal{A}_{0,k+1}), \\
&\quad \vdots \\
&\mathcal{B}_K = \bigcup_{k=1}^{K-2} (\mathcal{A}_{k,k+1} \setminus \mathcal{A}_{k,K+1}) \cup (\mathcal{A}_{0,K} \setminus \mathcal{A}_{0,K+1}).
\end{aligned} \tag{5-44}$$

$\mathcal{B}_1$ to $\mathcal{B}_K$ can be inferred by applying the algorithms in Table 5-1 to 5-3.

In comparison, because $W_{0,K+1} \preceq W_{0,1}$, in the CPMR-B protocol, the message $\boldsymbol{m}_0^{(t)}$ of source node 0 during time block t is only divided into two independent and disjoint sub message bit vectors $\boldsymbol{m}_{01}^{(t)} = \boldsymbol{m}_0^{(t)}(\mathcal{A}_{0,1} \setminus \mathcal{A}_{0,K+1})$ (the message bit vector which is in the set $\mathcal{A}_{0,1}$ but not in the set $\mathcal{A}_{0,K+1}$ of $\boldsymbol{m}_0^{(t)}$) and $\boldsymbol{m}_{0K+1}^{(t)} = \boldsymbol{m}_0^{(t)}(\mathcal{A}_{0,K+1})$ (the message bit vector which is in the set $\mathcal{A}_{0,K+1}$ of $\boldsymbol{m}_0^{(t)}$). At the same time, let $\boldsymbol{m}_k^{(t+l)}$ be the partial message delivered by relay node k during time block $t+l$, and $\boldsymbol{m}_k^{(t+l)}$ respectively be the partial message of the received message $\boldsymbol{m}_{k-1}^{(t+l-1)}$ from node $k-1$. $\boldsymbol{m}_k^{(t+l)}$ can be computed from the following equation:

$$\boldsymbol{m}_k^{(t+l)} = f_{\mathcal{C}_k}(\boldsymbol{m}_{k-1}^{(t+l-1)}), \tag{5-45}$$

where $f(\cdot)$ is the partial message function with the information sets of the source message $\boldsymbol{m}_{k-1}^{(t+l-1)}$ and the partial message as the parameters, in which, for an arbitrary $1 \le k \le K$

$$\mathcal{C}_k = (\mathcal{A}_{k-1,l} \setminus \mathcal{A}_{k-1,K+1}). \tag{5-46}$$

$\mathcal{C}_k$ can be inferred by the same method used to obtain $\mathcal{B}_1$ to $\mathcal{B}_K$.

5.5.2 Polar Encoding and Decoding Process for the Degraded MRN-ORCs

The encoding and decoding for the two protocols in the SDMRN-ORCs and the DDMRN-ORCs model are executed sequentially:

(1) Time block *t*

The source message vector $\boldsymbol{m}_0$ is firstly encoded into the codeword $\boldsymbol{x}_0$ and then is transmitted to relay node 1. Provided that the frozen set $\mathcal{F}_0$ is already known to relay node 1, relay node 1 can execute the SC decoding algorithm and thus reliably reconstruct the source message, when the transmission rate of the source message is:

$$\lim_{N\to\infty} R_0 = \lim_{N\to\infty} \frac{|\mathcal{A}_{0,1}|}{N} = I(X_0;Y_1). \tag{5-47}$$

On the other hand, although in the CPMR-A protocol, the remaining $K+1$ nodes receive the codeword $\boldsymbol{x}_0$ at the same time, node *K* only receives $NI(X_0;Y_{0,k})$ bits of information but still lacks $N(I(X_0;Y_1)-I(X_0;Y_{0,k}))$ bits of additional information about the sub message $\boldsymbol{m}_{01},\ldots\boldsymbol{m}_{0K}$ in $\boldsymbol{x}_0$ in order to fully recover the source message. Therefore, Node *k* only stores the received $\boldsymbol{x}_0$ and does not conduct further decoding until it has received sufficient helpful messages. However, in the CPMR-B protocol, only relay node 1 and the destination node receive the source message while other relay nodes do not receive the source message due to the long distance to the source node.

(2) Time block $t+l$

Because in the CPMR-A protocol $\mathcal{A}_{0,K+1} \subseteq \mathcal{A}_{0,K} \subseteq \cdots \subseteq \mathcal{A}_{0,1}$, relay node k in each time block knows exactly what partial message is required by relay node $k+1$ to reliably recover the source message. Therefore, after the partial message $\boldsymbol{m}_k$ of $\boldsymbol{m}_0^{(t)}$ is decoded by relay node k, it is forwarded to relay node $k+1$ and destination node $K+1$ to resolve the uncertainty about the source message $\boldsymbol{m}_0^{(t)}$ for relay node $k+1$ and destination node $K+1$, with the transmission rate being

$$R_k = \frac{\bigcup_{k=1}^{K-2}(|\mathcal{A}_{k,k+1}|-|\mathcal{A}_{k,K}|)+|\mathcal{A}_{0,k}|-|\mathcal{A}_{0,k+1}|}{N}. \tag{5-48}$$

With the help of the received $\boldsymbol{m}_k^{(t+l)} = f_{\mathcal{B}_k}(\boldsymbol{m}_0^{(t)})$, relay node $k+1$ can decode the source message $\boldsymbol{m}_0^{(t)}$. In the CPMR-B protocol, after relay node k decodes the received $\boldsymbol{m}_{k-1}$, it sends the partial message $\boldsymbol{m}_k$ to relay node $k+1$ and destination node $K+1$ to just resolve the destination node's uncertainty about the source message, with the transmission rate being:

$$R_1 = \frac{|\mathcal{A}_{0,1}|-|\mathcal{A}_{0,K+1}|}{N},$$

$$R_2 = \frac{|\mathcal{A}_{0,1}|-|\mathcal{A}_{0,K+1}|-|\mathcal{A}_{1,K+1}|}{N},$$

$$\vdots$$

$$R_k = \frac{|\mathcal{A}_{0,1}|-\bigcup_{k=1}^{K-1}(|\mathcal{A}_{k-1,K+1}|)}{N}. \tag{5-49}$$

With the help of $\boldsymbol{m}_k^{(t+l)} = f_{\mathcal{C}_k}(\boldsymbol{m}_{k-1}^{(t+l-1)})$, relay node $k+1$ can also correctly decode the message $\boldsymbol{m}_{k-1}^{(t+l-1)}$. At this point, although the destination node has already received

$\sum_{k=1}^{K-1} I(X_{k-1};Y_{k-1,K+1})$ bits of information during time block $t+l$, it still needs $N(I(X_0;Y_1)-\sum_{k=1}^{K-1} I(X_{k-1};Y_{k-1,K+1}))$ bits of reliable information to reconstruct the source message.

(3) Time block $t+K$

Relay node K encodes the message $\boldsymbol{m}_K$ into the codeword $\boldsymbol{x}_K$ of block length N, which is transmitted to the destination node at the rate

$$R_K = \frac{\bigcup_{k=1}^{K-2}(|\mathcal{A}_{k,k+1}|-|\mathcal{A}_{k,K+1}|)+|\mathcal{A}_{0,K}|-|\mathcal{A}_{0,K+1}|}{N} \tag{5-50}$$

to finally resolve the uncertainty about the source message $\boldsymbol{m}_0^{(t)}$. Lastly, the destination node uses the received helpful message $\boldsymbol{m}_k^{(t+l)} = f_{\mathcal{B}_k}(\boldsymbol{m}_0^{(t)})$ as the a priori information so as to reconstruct the source message $\boldsymbol{m}_0^{(t)}$. At this point, the codeword $\boldsymbol{x}_0$ is considered as the codeword $\boldsymbol{x}'_0$ delivered from the source node to the destination node during time block t, so the rate accordingly decreases to

$$\lim_{N\to\infty} R'_0 = \lim_{N\to\infty} \frac{|\mathcal{A}_{0,K+1}|}{N} = I(X_0;Y_{0,K+1}). \tag{5-51}$$

In the CPMR-B protocol, relay node K sends the partial message $\boldsymbol{m}_K$ to destination node $K+1$ to ultimately resolve the uncertainty about the source message $\boldsymbol{m}_0^{(t)}$, with the transmission rate being:

$$R_K = \frac{|\mathcal{A}_{0,1}|-\bigcup_{k=1}^{K}(|\mathcal{A}_{k-1,K+1}|)}{N}. \tag{5-52}$$

Lastly, the destination node uses the received helpful message $\boldsymbol{m}_k^{(t+l)} = f_{\mathcal{C}_k}(\boldsymbol{m}_{k-1}^{(t+l-1)})$ as the a priori information so as to reconstruct the source message $\boldsymbol{m}_0^{(t)}$.

Theorem 5-2 (Capacity Achievability of Polar Codes for the SDMRN-ORCs and DDMRN-ORCs): Consider SDMRN-ORCs and DDMRN-ORCs. Any transmission rate $0 < R < C_M^{OA}$ or $0 < R < C_M^{OB}$ is said to be achievable if there exists a sequence of polar codes of block length N, such that the block error probability $P_e = \Pr(\hat{\boldsymbol{m}}_0 \neq \boldsymbol{m}_0)$ at the destination under the SC decoding algorithm is bounded by $P_e \leq O(2^{-N^\beta})$ for any $0 < \beta < 1/2$ and sufficiently large N.

5.5.3 Asymptotic Performance of the Block Error Probability

This subsection proves Theorem 5-2. Let $\mathcal{E}_0$ represent the error event which occurs when the destination node reconstructs the source message $\boldsymbol{m}_0^{(t)}$. Let $\mathcal{E}_k$, $\mathcal{E}_K$ and $\mathcal{E}_{K+1}$ respectively be the event $\{\hat{\boldsymbol{m}}_k \neq \boldsymbol{m}_k\}$ occurring at relay node $k+1$, the event $\{\hat{\boldsymbol{m}}_{K-1} \neq \boldsymbol{m}_{K-1}\}$ occurring when relay node K recovers the partial message $\boldsymbol{m}_K$ delivered by relay node $K-1$ and the event $\{\hat{\boldsymbol{m}}_K \neq \boldsymbol{m}_K\}$ occurring when the destination node recovers the partial message $\boldsymbol{m}_K$ delivered by relay node K. At the same time, define $\mathcal{E}_k^c$, $\mathcal{E}_K^c$ and $\mathcal{E}_{K+1}^c$ as the complementary events of the above events. According to the formula of total probability, the block error probability of the CPMR scheme can be iteratively computed as

$$\begin{aligned}
P_e &= \Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_k^c) \\
&= \Pr(\mathcal{E}_0 \bigcap (\mathcal{E}_{k+1} \bigcup \mathcal{E}_{k+1}^c) \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_k^c) \\
&\leq \Pr(\mathcal{E}_0 \bigcap \mathcal{E}_{k+1} \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_k^c) + \Pr(\mathcal{E}_0 \bigcap \mathcal{E}_{k+1}^c \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_k^c) \\
&= \Pr(\mathcal{E}_{k+1} \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_k^c) \Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_k^c, \mathcal{E}_{k+1}) \\
&\quad + \Pr(\mathcal{E}_{k+1}^c \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_k^c) \Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_k^c, \mathcal{E}_{k+1}^c) \\
&\leq \Pr(\mathcal{E}_{k+1} \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_k^c) + \Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_k^c, \mathcal{E}_{k+1}^c). \qquad (5\text{-}53)
\end{aligned}$$

The first term in Equation (5-53) indicates the decoding error probability for relay node $k+1$ when the decoding is correct from relay node 1 to relay node *k*. If a polar code sequence is transmitted from node *k* to node $k+1$ at rate $R_k \le I(X_k;Y_{k,k+1})$, then based on the properties of polar codes, we can easily get

$$\Pr(\mathcal{E}_{k+1} \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_k^c) \le O(2^{-N^\beta}). \tag{5-54}$$

If destination node $K+1$ can correctly decode the partial message delivered from node *k*, the error probability for the destination node to reconstruct the source message can be written as:

$$\begin{aligned}\Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_K^c) &\le \Pr(\mathcal{E}_{K+1} \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_K^c)\Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_K^c, \mathcal{E}_{K+1}) \\ &\quad + \Pr(\mathcal{E}_{K+1}^c \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_K^c)\Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_K^c, \mathcal{E}_{K+1}^c) \\ &\le \Pr(\mathcal{E}_{K+1} \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_K^c) + \Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_K^c, \mathcal{E}_{K+1}^c). \end{aligned} \tag{5-55}$$

For any $R_K \le I(X_K;Y_{K,K+1})$, it satisfies

$$\Pr(\mathcal{E}_{K+1} \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_K^c) \le O(2^{-N^\beta}). \tag{5-56}$$

The second term in Equation (5-55) indicates the error probability for the destination node to reconstruct the source message $\boldsymbol{m}_0$ when the destination node receives the correct helpful partial messages from all the relays. Therefore, for any $R'_0 < I(X_0;Y_{0,K+1})$, its error probability is also upper bounded by

$$\Pr(\mathcal{E}_0 \mid \mathcal{E}_1^c, \mathcal{E}_2^c, \ldots, \mathcal{E}_K^c, \mathcal{E}_{K+1}^c) \le O(2^{-N^\beta}). \tag{5-57}$$

Therefore, when $N \to \infty$, the probability of the error event $\mathcal{E}_0$ is upper bounded by $O(2^{-N^\beta})$. □

As can be seen from the asymptotic performance analysis of the CPMR scheme of infinite block lengths in Theorem 5-1 and Theorem 5-2, the proposed scheme can achieve the capacities of SDMRN-ORCs and DDMRN-ORCs with arbitrarily small

error probabilities while the complexity of the encoder and decoder in the system is only $O(N \log N)$. Next, we will evaluate the actual performance of the CPMR scheme of finite block lengths and the feasibility of such a scheme.

5.6 Simulation Results and Analysis

For the convenience of evaluating the feasibility of the actual execution of the CPMR scheme and analyzing its performance, we carry out simulation based on the DDTRN-ORCs. The specific simulation parameters and their descriptions are as follows:

- Suppose the channel $W_{0,1}$ and the channel $W_{0,3}$ are BSC channels independent of each other and their transition probabilities are $W_{0,1}(y_1 \mid x_0) = 0.03$ and $W_{0,3}(y_{0,3} \mid x_0) = 0.12$;
- Suppose the channel $W_{1,2}$ and the channel $W_{1,3}$ are also BSC channels independent of each other and their transition probabilities are $W_{1,2}(y_2 \mid x_1) = 0.001$ and $W_{1,3}(y_{1,3} \mid x_1) = 0.005$;
- The discussion on the channel $W_{2,3}$ is based on two situations: One situation is that $W_{2,3}$ is a completely error-free channel with fixed channel capacity while the other situation is that $W_{2,3}$ is an independent BSC channel with a transition probability probability of $W_{2,3}(y_{2,3} \mid x_2) = 0.15$.

Based on the above suppositions, in the expression on the right-hand side of the equality in Equation (5-27), $I(W_{0,1}) = 0.8056$, $I(W_{0,3}) = 0.4706$, $I(W_{1,2}) = 0.9886$, $I(W_{1,3}) = 0.9546$, and $I(W_{2,3}) = 0.3092$. Therefore, the capacity of the DDTRN-ORCs is equal to the capacity of the channel $W_{0,1}$, i.e. $C_T^{OB} = I(W_{0,1}) = 0.8056$.

Fig. 5-10 illustrates the BER ($\Pr(\hat{\boldsymbol{m}}_0 \neq \boldsymbol{m}_0)$) curve of the CPMR scheme when the channel $W_{2,3}$ is a completely error-free channel (assume $I(W_{2,3}) = R_2$) versus varying channel transmission rate R_0 , transmission rate R_2 of relay node 2 and block length $N = 2^n$. As expected, with the increase of the block length and decrease of rate R_0 , BER is becoming smaller and smaller, which shows that the proposed scheme can achieve the channel capacity through proper design. In addition, by increasing the rate R_2 , BER also decreases accordingly. This is because the channel $W_{2,3}$ is a completely error-free channel while both the channel $W_{1,2}$ and the channel $W_{1,3}$ are also near error-free, which shows that almost the destination node can correctly decode all the partial messages delivered from relay node 1 and relay node 2 and the more reliable the partial messages forwarded by the relay nodes are, the more capable the destination node is of obtaining such helpful information in order to resolve the uncertainty about $\boldsymbol{m}_0$. When R_2 is increasing, it means that the quantity of messages relay node 2 assists in forwarding is also increasing and the number of messages the source-destination channel has to decode is decreasing, resulting in significant improvement in the performance.

Fig. 5-11 shows the BER curve when the channel $W_{2,3}$ is a BSC channel. As can be seen from the diagram, in this case, increasing R_2 does not necessarily improve the system performance. This situation occurs because some bits of the partial message forwarded by relay node 2 are reversed during transmission in the channel, which subsequently results in error bits. From a certain perspective, the occurrence of such error bits causes error floors in the system performance. At this point, the destination node cannot correctly recover the partial message forwarded by relay node 2 as the a priori information for the decoding of $\boldsymbol{m}_0$. Whether this message is decoded correctly

has an absolute impact on the performance in the decoding of $\boldsymbol{m}_0$, which means that the error event $\mathcal{E}_3$ in Equation (5-40) plays a crucial role on the performance of the whole system.

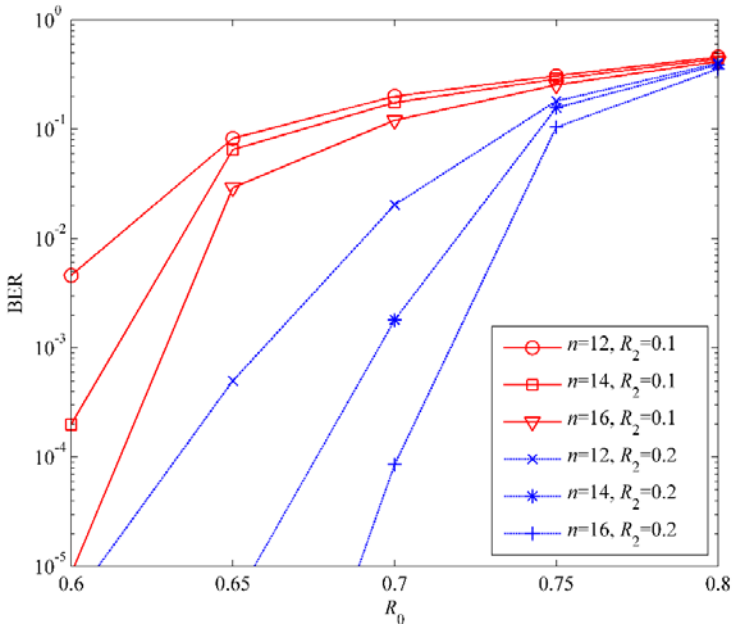


Fig. 5-10 Performance of the CPMR scheme for DDTRN-ORCs: $W_{2,3}$ is an error-free channel

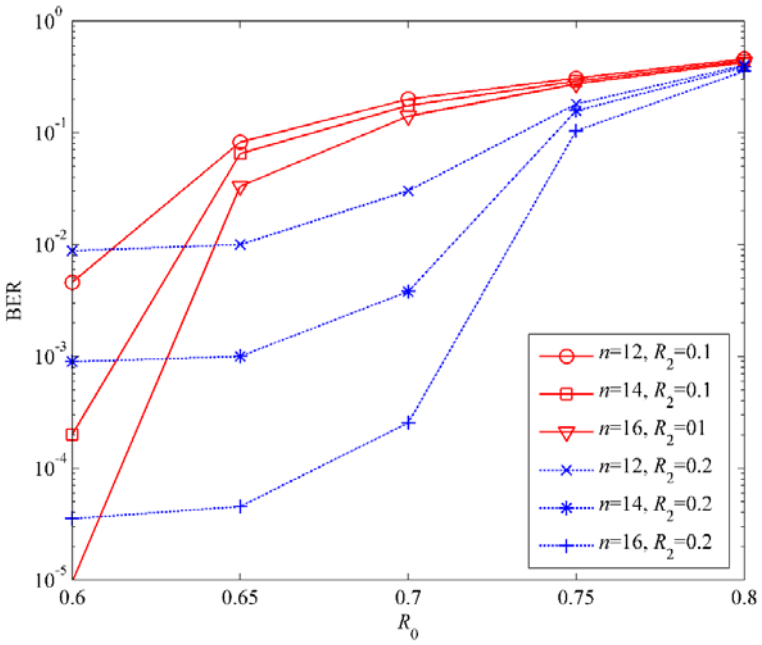


Fig. 5-11 Performance of the CPMR scheme for DDTRN-ORCs: $W_{2,3}$ is a BSC channel

To sum up, although the CPMR scheme for multiple-relay transmission systems is theoretically optimal, the simulation results show that, as a common problem with

the construction of polar codes of finite block lengths, the convergence speed of the CPMR system of finite block lengths is relatively slow. However, the application of the CPMR scheme in multiple-relay systems presents a specific and practical constructive channel coding method, which systematically realizes low encoding and decoding complexity of just $O(N \log N)$. As for the insufficiency in performance, the simulation results can be improved by the use of the SCLD decoding algorithm, which is expected to realize a performance approaching the relay capacity limit.

5.7 Summary

In this chapter, we introduced a novel DF strategy based on polar codes - the CPMR cooperative transmission scheme. The advantage of this scheme is that it can address the relatively high encoding and decoding complexity of non-constructive random coding methods in multiple-relay networks. By giving the relay node the flexibility of forwarding partial messages from the source or other relay nodes, the CPMR scheme of low complexity can be extended to multiple-relay networks. We explained the core idea of the CPMR schemes for two types of two-relay network models and then extended it to multiple-relay networks. In this chapter, we not only described the correlation between partial messages and the corresponding information sets but also analyzed the encoding and decoding process of constructive polar codes. The results show that CPMR schemes with low complexity can replace random coding methods with high complexity and can achieve the system capacity of these two types of degraded multiple-relay networks. Lastly, we evaluated the feasibility of the CPMR scheme through simulations.

Conclusion

As one of the key areas of research on wireless communications, cooperative communication technologies constitute an integral part of future broadband wireless communication systems and the study on such technologies has posed itself as a novel and challenging topic. Distributed channel coding for cooperative communication systems is a channel coding technique designed for decreasing the bit error rate for transmitted symbols and improving the system reliability. Thus, a proper distributed channel coding scheme is of great importance. Although common distributed Turbo codes and distributed LDPC codes can achieve a performance approaching the channel capacity limit, LDPC codes are non-systematic codes with high encoding complexity while Turbo codes have high decoding complexity. In view of that, in this thesis, we conducted research and analysis using the latest polar codes as our research objects. Compared with Turbo codes and LDPC codes, polar codes benefit from fast encoding and parallel decoding. Therefore, research on the application of polar codes in wireless networks is of great value. With a comprehensive knowledge of the current development and research on cooperative communication technologies and distributed channel coding both at home and abroad, we carried out an in-depth study on some key technologies for polar codes and proposed new ideas and methods regarding the improvement of performance of polar codes for wireless communication channels and the realization of polar codes in half-duplex single-relay channels and multiple-relay networks, which led to some valuable research results. Our findings can be summarized into the following three points:

Firstly, in response to the slow convergence speed of polar codes and the high error floors of LDGM codes, by utilizing the idea of concatenated codes, we proposed a SCPL coding scheme with low complexity which uses polar codes as the outer codes

and LDGM codes as the inner codes. By constructing SCPL codes and deriving the corresponding message iterative decoding algorithm based on the Tanner graph, we improved the basic theoretical analysis of SCPL codes. Through actual simulations of SCPL codes, we properly selected SCPL coding design parameters to evaluate the BER performance under different parameters. Our research findings show that SCPL codes can effectively address the problems in practical applications of polar codes. They not only overcome the error floor problem but are also easier to be implemented into practical applications while maintaining a relatively low coding complexity. In the meantime, they also guarantee a performance approaching the Shannon theoretical capacity limit. They represent a new pioneering direction in research on polar codes for areas such as deep space communications, long-distance optical fiber communications and data storage.

Then, we examined the challenges for polar codes in cooperative relay transmission systems in terms of capacity achievability, difficulty in realization, design of cooperative transmission protocols and actual performance. Based on the degraded half-duplex single-relay channel model, we proposed a CPMR transmission scheme with low complexity. Through a detailed theoretical study on the polar coding method and the CPMR protocol, it was shown that, while maintaining the low complexity of polar coding, the CPMR scheme can asymptotically achieve the capacity limit of the degraded half-duplex single-relay channel. The upper bound on the average block error probability under the SC decoding algorithm was also derived. Regarding the application of the CPMR scheme in the actual model, we, employing “Turbo” theory, designed a JISPIC receiver structure based on polar codes using MAP detection on the receiver side in order to recover the two streams of superposed information from the source node and the relay node in the MAC phase. Through the analysis of the relay

channel capacity limit, we explained the key factors affecting the capacity limit of the channel model, based on which, we conducted simulations of the CPMR scheme of finite block lengths using BPSK modulation for the AWGN channel. Compared with the already mature LDPC codes, the proposed scheme inherits the advantages of low encoding and decoding complexity of polar codes, achieves a superior performance, and effectively solves the problems with the actual application of polar codes in the DF relay channel.

Lastly, to address the high encoding complexity of non-constructive coding methods in multiple-relay systems, we gave the relay node the flexibility of forwarding partial messages from the source or other relay nodes so that the CPMR scheme with low complexity can be extended to multiple-relay networks. Regarding the two types of degraded multiple-relay networks with orthogonal receiver components, we described the correlation between partial messages and corresponding information sets for CPMR protocols, analyzed the constructive encoding and decoding process of polar codes, proved the asymptotic achievability of the capacity of the CPMR scheme, and through equation derivation, we found that the upper bound on the block error probability inherits that of the polar codes. Lastly, we evaluated the feasibility of the CPMR scheme in multiple-relay transmission systems through simulations.

In this thesis, we conducted an in-depth study on the application of polar codes in wireless communications with emphasis on the analysis of some specific methods for implementing polar codes in relay systems, from which some meaningful conclusions are drawn. However, due to the limitation of time and resources, it was not possible to research on all the technologies for polar codes and therefore our current work can still be further improved and expanded mainly in the following two areas:

Firstly, the SPCL scheme proposed in this thesis, which is a near-optimal solution,

was obtained through computer simulations, and we did not provide any theoretical analysis regarding how to search for the optimal combination of outer and inner coding rates and the best inner code weight. Provided that there is a solid theoretical foundation regarding polar codes, the next step will be to research on the optimum SCPL coding scheme with the aim of achieving the optimum system performance.

Secondly, as we focus on the theoretical analysis of coding schemes and communication protocols based on polar codes in relay systems, for the convenience of explaining the basics of the application of the CPMR scheme in multiple-relay systems and evaluating its performance, we only paid attention to two types of degraded multiple-relay network models with orthogonal receiver components in Chapter 5. The application of the CPMR scheme in general multiple-relay networks would necessitate a background knowledge of source coding. Therefore, one of the future tasks would be to incorporate the source coding technology in the design of a more general coding strategy with the aim of applying the CPMR scheme in a wider range of general multiple-relay network topologies.

References

[1] Goldsmith A, Jafar S, Jindal N, et al. Capacity Limits of MIMO Channels[J]. IEEE Journal on Selected Areas in Communications, 2003, 21(5):684-702.

[2] Catreux S, Greenstein L J and Erceg V. Some Results and Insights on the Performance Gains of MIMO Systems[J]. IEEE Journal on Selected Areas in Communications, 2003, 21(5):839-847.

[3] Narula A, Trott M D and Wornell GW. Performance Limits of Coded Diversity Methods for Transmitter Antenna Arrays[J]. IEEE Transactions on Information Theory, 1999, 45(7): 2418–2433.

[4] Sendonaris A, Erkip E, Aazhang B. User Cooperation Diversity-Part I: System Description[J]. IEEE Transactions on Communications, 2003, 51(11): 1927-1938.

[5] Sendonaris A, Erkip E, Aazhang B. User Cooperation Diversity-Part II: Implementation Aspects and Performance Analysis[J]. IEEE Transactions on Communications, 2003, 51(11): 1939-1948.

[6] Chen D and Laneman J N. Modulation and Demodulation for Cooperative Diversity in Wireless Systems[J]. IEEE Transactions on Wireless Communications, 2006, 5(7): 1785-1794.

[7] Kramer G, Gastpar M and Gupta P. Cooperative Strategies and Capacity Theorems for Relay Networks[J]. IEEE Transactions on Information Theory, 2005, 51(9): 3037-3063.

[8] Larsson E G and Vojcic B R. Cooperative Transmit Diversity Based on Superposition Modulation[J]. IEEE Communications letters, 2005, 9(9): 778–780.

[9] Laneman J N, Tse D N C and Wornell G W. Cooperative Diversity in Wireless

Networks: Efficient Protocols and Outage Behavior[J]. IEEE Transactions on Information Theory, 2004, 50(12): 3062-3080.

[10] Sendonaris A, Erkip E and Aazhang B. Increasing Uplink Capacity via User Cooperation Diversity[C]. 1998 IEEE International Symposium on Information Theory, Cambridge, 1998: 16-21.

[11] Nosratinia A, Hunter T E and Hedayat A. Cooperative Communication in Wireless Network[J]. IEEE Communication Magazine, 2004, 42(10): 74-80.

[12] Kramer G, Maric I and Yates R D. Foundations and Trends in Networking: Cooperative Communications[J]. NOW publishers, 2007.

[13] Van der Meulen E C. Three-Terminal Communication Channels[J]. Advanced Applied Probability, 1971, 3: 120-154.

[14] Janani M, Hedayat A, Hunter T E, et al. Coded Cooperation in Wireless Communications: Space-Time Transmission and Iterative Decoding[J]. IEEE Trans. Signal Processing, 2004, 52(2): 362-371.

[15] Zhao B and Valenti M C. Distributed Turbo Coded Diversity for Relay Channel[J]. Electronics Letters, 2003, 39(10): 786-787.

[16] Zhang Z and Duman T M. Capacity Approaching Turbo Coding and Iterative Decoding for Relay Channels[J]. IEEE Transactions Communications, 2005, 53(11): 1895-1905.

[17] Zhang Z and Duman T M. Capacity Approaching Turbo Coding for Half Duplex Relaying[J]. IEEE Transactions Communications, 2007, 55(10): 1895-1906.

[18] Chakrabarti A, Baynast A, Sabharwal A, et al. Low Density Parity Check Codes for the Relay Channel[J]. IEEE Journal on Selected Areas in Communications, 2007, 25(2): 280-291.

[19] Hu J and Duman T M. Low Density Parity Check Codes over Wireless Relay Channels[J]. IEEE Transactions on Wireless Communications, 2007, 6(9): 3384-3394.

[20] Razaghi P and Yu W. Bilayer Low-Density Parity-Check Codes for Decode-and-Forward in Relay Channel[J]. IEEE Transactions on Information Theory, 2007, 53(10): 3723-3739.

[21] Shannon C E. A Mathematical Theory of Communication[J]. Bell System Technical Journal, 1948, 27(3): 379-423.

[22] Verdu S. Fifty Years of Shannon Theory[J]. IEEE Transactions on Information Theory, 1998, 44(6): 2057-2078.

[23] Proakis J and Salehi M. Digital Communications, 5th Edition[M]. McGraw-Hill Science/Engineering/Math, 2007.

[24] Sklar B. Digital Communications: Fundamentals and Applications, 2nd Edition[M]. Prentice Hall, 2001.

[25] Wang X M, Xiao G Z. Error Correction Codes: Principles and Methods (Modified version)[M]. Xidian University Press, 2001.

[26] Costello D J and Forney G D. Channel Coding: The Road to Channel Capacity[J]. Proceedings of The IEEE, 2007, 95(6): 1150-1177.

[27] Lin S and Costello D J. Error Control Coding: Fundamentals and Applications, 2nd Edition[M]. Prentice-Hall, 2004.

[28] Gallager R G. Low Density Parity-Check Codes[D]. Cambridge, MA: MIT Press, 1963.

[29] Gallager R G. Low Density Parity-Check Codes[J]. IRE Transactions on Information Theory, 1962, 8(1): 21-28.

[30] Viterbi A J and Omura J K Principles of Digital Communication and

Coding[M]. Dover Publications, 2009.

[31] Golay M J E. Notes on Digital Coding, Proceedings of The IEEE, 1949, 37: 657.

[32] Hamming R W. Error Detecting and Error Correcting Codes[J]. Bell System Technical Journal, 1950, 26(2): 147-160.

[33] Reed I S. A Class of Multiple-Error-Correcting Codes and Their Decoding Scheme[J]. IRE Transactions on Information Theory, 1954, 4(4): 38-42.

[34] Muller D E. Applications of Boolean Algebra to Switching Circuit Design and to Error Detection[J]. IRE Transactions on Electronic Computers, 1954, 3: 6-12.

[35] Blahut R E. Algebraic Codes for Data Transmission, 1st Edition[M]. Cambridge University Press, 2003.

[36] Roth R. Introduction to Coding Theory[M]. Cambridge University Press, 2006.

[37] Elias P. Coding for Noisy Channels[J]. IRE Convention Record, 1955, 4:37-46.

[38] Fano R. A Heuristic Ciscussion of Probabilistic Decoding[J] IEEE Transactions on Information Theory,1963, 9(2): 64-74.

[39] Wozencraft J M and Reiffen B. Sequential Decoding[M]. Cambridge, MA: MIT Press, 1961.

[40] Massey J L. Threshold Decoding[M]. Cambridge, MA: MIT Press, 1963.

[41] Sklar B. How I Learned to Love the Trellis[J]. IEEE Signal Processing Magazine, 2003, 20(3): 87-102.

[42] Johannesson R and Zigangirov K. Fundamentals of Convolutional Coding[M]. Orient Blackswan, 1999.

[43] Forney G. The Viterbi Algorithm[J]. Proceedings of the IEEE, 1973, 61(3):

268-278.

[44] Viterbi A. Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm[J]. IEEE transactions on Information Theory, 1967, 13(2): 260-269.

[45] Bahl L, Cocke J Jelinek F, et al. Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate[J]. IEEE Transactions on Information Theory, 1974, 20(2): 284-287.

[46] Forney G. Concatenated Codes[M]. Cambridge, MA: MIT Press, 1966.

[47] Wicker S B and Bhargava V K. Reed-Solomon Codes and Their Applications[M]. John Wiley & Sons, 1999.

[48] Berrou C, Glavieux A, and Thitimajshima P. Near Shannon Limit Error Correcting Coding and Decoding: Turbo Codes[C]. in Proceedings of the 1993 IEEE International Conference on Communication, Geneva, Switzerland, 1993: 1064-1070.

[49] Richardson T and Urbanke R. Modern Coding Theory[M]. Cambridge University Press, 2008.

[50] Ryan W and Lin S. Channel Codes: Classical and Modern[M]. Cambridge University Press, 2009.

[51] MacKay D J C. Good Error-Correcting Codes Based on Very Sparse Matrices[J]. IEEE Transactions on Information Theory, 1999, 45(2): 399-431.

[52] MacKay D J C and Neal R M. Near Shannon Limit Performance of Low-Density Parity Check Codes[J]. Electronics Letters, 1997, 33(6): 457-458.

[53] Chung S Y, Forney G, Richardson T, et al. On the Design of Low-Density Parity-Check Codes within 0.0045dB of the Shannon Limit. IEEE Communications Letters, 2001, 5(2):58-60.

[54] Richardson T J, Shokrollahi M A, and Urbanke R L. Design of Capacity-Approaching Irregular Low-Density Parity-Check Codes[J]. IEEE Transactions on Information Theory, 2001, 47(2):619-637.

[55] Chung S Y, Richardson T J and Urbanke R L. Analysis of Sum-Product Decoding of Low-Density Parity-Check Codes Using a Gaussian Approximation[J]. IEEE Transactions on Information Theory, 2001: 47(2):657-670.

[56] Garcia-Frias J and Zhong W. Approaching Shannon Performance by Iterative Decoding of Linear Codes with Low-Density Generator Matrix[J]. IEEE Communications Letters，2003, 7(6): 266-268.

[57] Zhong W and Garcia-Frias J. LDGM Codes for Channel Coding and Joint Source-Channel Coding of Correlated Sources[J]. EURASIP Journal on Applied Signal Processing, 2005, 6: 942-953.

[58] Kim J S and Song H Y. Concatenated LDGM Codes with Reduced Decoder Complexity[C]. 2006 IEEE Vehicular Technology Conference, 2006: 1363-1366.

[59] Kumano Y and Ohtsuki T. On Construction of Concatenated Low-Density Generator Matrix Codes[C]. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, 2007: 442-445.

[60] Vazquez-Araujo F J, Gonzalez-Lopez M, Castedo L, et al. Serially-Concatenated LDGM Code for MIMO Channels[J]. IEEE Transactions on Communications, 2007, 6(8): 2860-2871.

[61] Ryan W E. Concatenated Convolutional Codes and Iterative Decoding[M]. in Wiley Encyclopedia of Telecommunications, John Wiley & Sons, Inc., 2003.

[62] Shokrollahi A. LDPC Codes: An Introduction[R]. Digital Fountain, Inc., Tech.

Rep, 2003.

[63] Arikan E. Channel Polarization: A Method for Constructing Capacity Achieving Codes for Symmetric Binary-Input Memoryless Channels[J]. IEEE Transactions on Information Theory, 2009, 55(7): 3051-3073.

[64] Arikan E and Telatar E. On the Rate of Channel Polarization[C]. In: 2008 International Symposium on Information Theory, Seoul, Korea, 2008: 1493-1495.

[65] Korada S B. Polar Codes for Channel and Source Coding[D]. EPFL, Lausanne, 2009.

[66] Sasoglu E, Telatar E and Arikan E. Polarization for Arbitrary Discrete Memoryless Channels[C]. Proceedings of the 2009 IEEE Information Theory Workshop on Networking and Information Theory (ITW'09), Volos, Greece, 2009: 144-148.

[67] Hassani S H and Urbanke R. On the Scaling of Polar Codes: I. The Behavior of Polarized Channels[C]. Proceedings in 2010 IEEE International Symposium on Information Theory, Austin, TX, 2010: 874-878.

[68] Tanaka T and Mori R. Refined Rate of Channel Polarization[C]. Proceedings in 2010 IEEE International Symposium on Information Theory, Austin, TX, 2010: 889-893.

[69] Hassani S H, Mori R, Tanaka T, et al. Rate-Dependent Analysis of the Asymptotic Behavior of Channel Polarization[J]. IEEE Transactions on Information Theory, 2013, 59(4): 2267-2276.

[70] Korada S B and Sasoglu E and Urbanke R. Polar Codes: Characterization of Exponent, Bounds, and Constructions[J]. IEEE Transactions on Information Theory, 2010, 56(12): 6253-6264.

[71] Tal I and Vardy A. How to Construct Polar Codes[J]. IEEE Transactions on Information Theory, 2012, submitted for publication.

[72] Tal I and Vardy A. List Decoding of Polar Codes[DB/OL]. (2012-05-31) [2012-12-21]. http://arxiv.org/abs/1206.0050

[73] Hassani S H, Korada S B and Urbanke R. The Compound Capacity of Polar Codes[C]. Proceedings of the Forty-Seventh Annual Allerton Conference on Communication, Control and Computing, Allerton, Monticello, Illinois, USA, 2009: 16-21.

[74] Naveen G, Emmanuelle A. Polar Codes for the Deterministic Broadcast Channel[C]. Proceedings of the 2012 International Zurich Seminar on Communications International Zurich Seminar on Communications, Zurich, Switzerland, 2012.

[75] Sasoglu E, Telatar E, and Yeh E. Polar Codes for the Two User Multiple-Access Channel[DB/OL]. (2010-06-22) [2012-12-21]. http://arxiv.org/abs/1006.4255.

[76] Emmanuel A and Telatar E. Polar Codes for the *m*-User Multiple Access Channel[J]. IEEE Transactions on Information Theory, 2012, 58(8): 5437-5448.

[77] Cover T M and Gamal E. Capacity Theorems for the Relay Channel[J]. IEEE Transactions on Information Theory, 1979, 25(5):572-584.

[78] Laneman J N. Cooperative Diversity in Wireless Networks: Algorithms and Architectures[D]. MIT, Cambridge, MA, 2002.

[79] Pabst R, Walke B H, Schultz D C, et al. Relay-Based Deployment Concepts for Wireless and Mobile Broadband Radio[J]. IEEE Communications Magazine, 2004, 42(9): 80-89.

[80] King R C. Multiple Access Channels with Generalized Feedback[D]. Stanford University., Stanford, CA, 1978.

[81] Slepian D and Wolf J K. A Coding Theorem for Multiple Access Channels with Correlated Sources[J]. Bell System Technical Journal, 1973, 52:1037-1076.

[82] Gaarder N T and Wolf J K. The Capacity Region of a Multiple-Access Discrete Memoryless Channel can Increase with Feedback[J]. IEEE Transactions on Information Theory, 1975, 21(1): 100-102.

[83] Cover T M and Leung C S K. An Achievable Rate Region for the Multiple-Access Channel with Feedback[J]. IEEE Transactions on Information Theory, 1981, 27(3): 292-298.

[84] Carleial A B. Multiple-Access Channels with Different Generalized Feedback Signals. IEEE Transactions on Information Theory, 1982, 28(6): 841-850. 1982.

[85] Willems F M G. Information Theoretical Results for the Discrete Memoryless Multiple Access Channel[D]. Katholieke University, Leuven, Belgium, 1982.

[86] Zeng C M, Kuhlmann F and Buzo A. Achievability Proof of Some Multiuser Channel Coding Theorems Using Backward Decoding[J]. IEEE Transactions on Information Theory, 1989, 35(6): 1160-1165.

[87] Laneman J N and Kramer G. Window Decoding for the Multi-Access Channel with Generalized Feedback[C]. IEEE International Symposium on Information Theory, Chicago, IL, 2004: 281.

[88] Sankaranarayanan L, Kramer G and Mandayam N B. Capacity Theorems for the Multiple-Access Relay Channel[C]. 42nd Annual Allerton Conference on Communications, Control, and Computing, Monticello, IL, 2004: 1782-1791.

[89] Aref M R. Information Flow in Relay Networks[D]. Stanford University, Stanford, CA, 1980.

[90] Gamal E. On Information Flow in Relay Networks[C]. IEEE National Telecommunications Conference, Miami, FL, 1981: D4.1.1-D4.1.4.

[91] Cover T M and Thomas J A. Elements of Information Theory[M]. New York: Wiley, 1991.

[92] Gupta P and Kumar P R. Toward an Information Theory of Large Networks: An Achievable Rate Region[J]. IEEE Transactions on Information Theory, 2003, 49(8): 1877-1894.

[93] Xie L L and Kumar P R. A Network Information Theory for Wireless Communication: Scaling Laws and Optimal Operation[J]. IEEE Transactions on Information Theory, 2004, 50(5): 748-767.

[94] Xie L L and Kumar P R. An Achievable Rate for the Multiple-Level Relay Channel[J]. IEEE Transactions on Information Theory, 2005, 51(4): 1348-1358.

[95] Kramer G, Gastpar M and Gupta P. Capacity Theorems for Wreless Relay Channels[C]. 41st Annual Allerton Conference on Communications, Control, and Computing, Monticello, IL, 2003: 1074-1083.

[96] Razaghi P and Yu W. Parity Forwarding for Multiple-Relay Networks[J]. IEEE Transactions on Information Theory, 2009, 55(1): 158-173.

[97] Hunter T E and Nosratinia A. Cooperative Diversity through Coding[C]. IEEE International Symposium on Information Theory, Lausanne, Switzerland, 2002: 220.

[98] Nabar R U, Bolcskei H and Kneubuhler F W. Fading Relay Channels: Performance Limits and Space-Time Signal Design[J] IEEE Journal on Selected Areas in Communications, 2004, 22(6): 1099-1109.

[99] Stefanov A and Erkip E. Cooperative Space-Time Coding for Wireless Networks[J]. IEEE Transactions on Communications, 2005, 53(11): 1804-1809.

[100] Janani M, Hedayat A, Hunter T et al. Coded Cooperation in Wireless Communications: Space-Time Transmission and Iterative Decoding[J]. IEEE Transactions on Signal Processing, 2004, 52(2): 362-371.

[101] Souryal M R and Vojcic B R. Cooperative Turbo Coding with Time Varying Rayleigh Fading Channels[C]. International Conference on Communications, Paris, France, 2004: 356-360.

[102] Li Y, Vucetic B, Wong T F, et al. Distributed Turbo Coding With Soft Information Relaying in Multihop Relay Networks[J]. IEEE Journal on Selected Areas in Communications, 2006, 24(11): 2040-2050.

[103] Chakrabarti A, Baynast A, Sabharwal A, et al. LDPC Code Design for Half-Duplex Decode-and-Forward Relaying[C]. 43rd Annual Allerton Conference on Communications, Control, and Computing, Allerton, IL, 2005: 972-981.

[104] Li C, Yue G, Khojastepour M A, et al. LDPC-Coded Cooperative Relay Systems: Performance Analysis and Code Design[J]. IEEE Transactions on Communications, 2007, 56(3): 485-496.

[105] Razaghi P and Yu W. Bilayer Low-Density Parity-Check Codes for Decode-and-Forward in Relay Channels[J]. IEEE Transactions on Information Theory, 2007, 53(10): 3723-3739.

[106] Cances J P and Meghdadi V. Optimized Low Density Parity Check Code Designs for Half Duplex Relay Channels[J]. IEEE Transactions on Wireless Communications, 2009, 8(7): 3390-3395.

[107] Andersson M, Rathi V, Thobaben R, et al. Nested Polar Codes for Wiretap and Relay Channels[J]. IEEE Communications Letters, 2010, 14(8): 752-754.

[108] Blasco-Serrano R, Thobaben R, Andersson M, et al. Polar Codes for Cooperative Relaying[J]. IEEE Transactions on Wireless Communications, 2012, 60(11): 3263-3273.

[109] Karzand M. Polar Codes for Degraded Relay Channels[C]. Proceedings of the 2012 International Zurich Seminar on Communications International Zurich Seminar on Communications, Zurich, Switzerland, 2012.

[110] Hussami N, Korada S B and Urbanke R. Performance of Polar Codes for Channel and Source Coding[DB/OL]. (2009-05-22) [2012-12-21]. http://arxiv.org/abs/0901.2370.

[111] Arikan E. A Performance Comparison of Polar Codes and Reed-Muller Codes[J]. IEEE Communications Letters, 2008, 12(6): 447-449.

[112] Mori R and Tanaka T. Performance of Polar Codes with Construction Using Density Evolution[J]. IEEE Communications Letters, 2009, 13(7): 519-521.

[113] Mori R and Tanaka T. Performance and Construction of Polar Codes on Symmetric Binary-Input Memoryless Channels[C]. 2009 International Symposium on Information Theory, Seoul, Korea, 2009: 1496-1500.

[114] Eslami A and Pishro-Nik H. On Bit Error Rate Performance of Polar Codes in Finite Regime[C].48th Annual Allerton Conference on Communication, Control, and Computing, Allerton, IL, 2010: 188-194.

[115] Eslami A and Pishro-Nik H. A Practical Approach to Polar Codes[C]. IEEE International Symposium on Information Theory, St. Petersburg, 2011: 16-20.

[116] Goela N, Korada S B and Gastpar M. On LP Decoding of Polar Codes[C]. 2010 IEEE Information Theory Workshop, Dublin, 2010: 1-5.

[117] Alamdar-Yazdi A and Kschischang F R. A Simplified Successive-Cancellation Decoder for Polar Codes[J]. IEEE Communications Letters, 2011, 15(12): 1378-1380.

[118] Arıkan E. Systematic Polar Coding[J]. IEEE Communications Letters, 2011, 15(8): 860-862.

[119] Trifonov P. Efficient Design and Decoding of Polar Codes[J]. Transactions on Communications, 2012, 60(11): 3221-3227.

[120] Chen K, Niu K and Lin J R. List Successive Cancellation Decoding of Polar Codes[J]. Electronics Letters, 2012, 48(9): 500-501.

[121] Niu K and Chen K. Stack Decoding of Polar Codes[J]. Electronics Letters, 2012, 48(12): 695-697.

[122] Moon T K. Error Correction Coding: Mathematical Methods and Algorithms, 1st Edition[M]. Wiley-Interscience, 2005.

[123] Gallager R G. Information Theory and Reliable Communication[M]. Wiley, 1968.

[124] Bakshi M, Jaggi S and Effros M. Concatenated Polar codes[C]. 2010 IEEE International Symposium on Information Theory, Austin, TX, 2010: 918-922.

[125] Lin S and Costello D J. Error Control Coding, 2nd Edtion[M]. Prentice Hall, 2004.

[126] Kschischang F R, Frey B J and Loeliger H A. Factor Graphs and the Sum-Product Algorithm[J] IEEE Transactions on Information Theory, 1998, 47: 498–519.

[127] Wiberg N, Loeliger H A and Kotter R. Codes and Iterative Decoding on General Graphs[J]. European Transactions on Telecommunications, 1995, 6(5): 513-525.

[128] Tanner R. A Recursive Approach to Low Complexity Codes[J] IEEE Transactions on Information Theory, 1981, 27(5): 533-547.

[129] Johnson S J. Iterative Error Correction: Turbo, Low-Density Parity-Check and Repeat-Accumulate Codes, 1st Edition[M]. Cambridge University Press, 2009.

[130] Richardson T J and Urbanke R L. The Capacity of Low-Density Parity-Check Codes under Message-Passing Decoding[J]. IEEE Transactions on Information Theory, 2001, 47(2): 599-618.

[131] Arikan E. Polar Codes: A Pipelined Implementation[C]. 4th International Symposium on Broadband Communications (ISBC2010), Melaka, Malaysia, 2010: 11-14.

[132] Forney G D. Codes on Graphs: Normal Realizations[J]. IEEE Transactions on Information Theory, 2001, 47(2): 520-548.

[133] Hu X Y, Eleftheriou E and Arnold D M. Progressive Edge-Growth Tanner Graphs[C]. in Proceedings of. IEEE GLOBECOM 2001, San Antonio, TX, 2001: 995-1001.

[134] Ashikhmin A E, Kramer G, Brink S T. Extrinsic Information Transfer Functions: Model and Erasure Channel Properties[J]. IEEE Transactions on Information Theory, 2004, 50(11): 2657-2673.

[135] Khojastepour M. Distributed Cooperative Communications in Wireless Networks[D]. Rice University, 2005.

[136] Wang X and Poor H V. Iterative (Turbo) Soft Interference Cancellation and Decoding for Coded CDMA[J] IEEE Transactions on Communications, 1999, 47(7): 1046-1061.

[137] Vucetic B and Yuan J. Turbo Codes: Principles and Applications[M]. Norwell, MA: Kluwer, 2000.

[138] Kim Y H. Coding Techniques for Primitive Relay Channels[C]. Annual Allerton Conference on Communication, Control and Computing, Allerton House, UIUC, Illinois, USA, 2007: 129-135.

List of Publications

1. Bin Duo, Zhenyong Wang and Xuemai Gu. Achieving the Capacity of Degraded Half-Duplex Relay Channel Using Polar Coding. Chinese Journal of Aeronautics, 27(3): 584-592, 2014.
2. Bin Duo, Peng Wang, Yonghui Li, Branka Vucetic. Secure transmission for relay-eavesdropper channels using polar coding. IEEE ICC 2014; 2197-2202.
3. Bin Duo, Zhenyong Wang, Xuemai Gu. A new achievable rate region for the relay channel with forwarding own messages. Information Technology Journal, 13(2): 359-364, 2014.
4. Bin Duo, Zhenyong Wang and Xuemai Gu. Cooperative Partial Message Relaying Based on Distributed Polar Codes for the Two-Relay Network. Applied Mechanics and Materials, 303-306, 1974-1983, 2013.
5. Bin Duo, Zhenyong Wang, Xuemai Gu. Achieving the secrecy capacity of relay-eavesdropper channel using polar codes. Information Technology Journal, 2013. (Accepted).
6. Bin Duo, Zhenyong Wang, Xuemai Gu. On construction of low complexity serially-concatenated LDGM codes. Journal of Harbin Institute of Technology, 45(5): 25-29, 2013.
7. Bin Duo, Zhenyong Wang, Mingchuan Yang, Xuemai Gu. Low complexity polar product coding for decode-and-forward cooperative relay channel. Chinese High Technology Letters, 23(2):153-159, 2013.
8. Bin Duo, Zhenyong Wang, Xuemai Gu. On construction of Shannon capacity-approaching low complexity polar-LDGM codes. Chinese High Technology Letters, 23(6): 564-570, 2013.
9. Bin Duo, Zhenyong Wang, Xuemai Gu. Serially-concatenated Polar-LDGM

codes for satellite communications systems. Submitted to Science China Information Sciences, 2015.

10. Bin Duo, Zhenyong Wang, Xuemai Gu. Nested polar codes for degraded multiple-relay networks. Submitted to Wireless Communications and Mobile Computing, 2015.