<table>
<thead>
<tr>
<th>Title</th>
<th>A Tutorial on Lossy Forwarding Cooperative Relaying</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>He, Jiguang; Tervo, Valtteri; Zhou, Xiaobo; He, Xin; Qian, Shen; Cheng, Meng; Juntti, Markku; Matsumoto, Tad</td>
</tr>
<tr>
<td>Citation</td>
<td>IEEE Communications Surveys &amp; Tutorials</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2018-08-23</td>
</tr>
<tr>
<td>Type</td>
<td>Journal Article</td>
</tr>
<tr>
<td>Text version</td>
<td>Publisher</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10119/15423">http://hdl.handle.net/10119/15423</a></td>
</tr>
<tr>
<td>Rights</td>
<td>Copyright (C) 2018 IEEE. IEEE Communications Surveys &amp; Tutorials, 2018, DOI:10.1109/COMST.2018.2866711. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See <a href="http://www.ieee.org/publications_standards/publications/rights/index.html">http://www.ieee.org/publications_standards/publications/rights/index.html</a> for more information.</td>
</tr>
<tr>
<td>Description</td>
<td></td>
</tr>
</tbody>
</table>
A Tutorial on Lossy Forwarding Cooperative Relaying

Jiguang He, Student Member, IEEE, Valtteri Tervo, Xiaobo Zhou, Member, IEEE, Xin He, Member, IEEE, Shen Qian, Student Member, IEEE, Meng Cheng, Markku Juntti, Senior Member, IEEE, and Tad Matsumoto, Fellow, IEEE

Abstract—Lossy decode-and-forward (DF) relaying, also referred to as lossy forwarding (LF), can significantly enhance the transmission reliability and expand the communication coverage at the cost of a small increase in computational effort compared to its DF counterpart. Furthermore, it can further simplify the operations at the relay nodes by removing the error-detecting operation, e.g., cyclic redundancy check, which is used in the conventional DF systems. Due to these advantages, LF has been intensively investigated with the aim of its applications to various cooperative communication networks with different topologies. This paper offers a comprehensive literature review on the LF relaying strategy and makes comparisons between LF and DF. Five basic exemplifying scenarios are taken into consideration. These are the three-node network, the single-source multi-relay network with direct source-to-destination link, the multiple access relay channel, the two-way relay network, and the general multi-source multi-relay network. The paper includes not only theoretical performance limit analyses, but also performance evaluation by employing low-complexity accumulator aided turbo codes at the sources and relays as well as joint decoding at the destination. As expected, the performance enhancement in terms of outage probability, frame error rate, and $\epsilon$-outage achievable rate by LF over DF is significant, which is demonstrated in all the exemplifying scenarios in the literatures. Hence LF has a great potential to be applied to future 5G wireless communication networks, e.g., device-to-device, vehicle-to-vehicle, and machine-type communications, which are composed of the aforementioned exemplifying scenarios.

Index Terms—Decode-and-forward, distributed turbo code, joint decoding, lossy forwarding, lossy source-channel separation theorem, multiple access relay channel, Slepian-Wolf, source coding with a helper.

I. INTRODUCTION

FUTURE wireless communications aim at achieving ultra-low latency, very low power consumption, low manufacturing costs, high reliability, extremely high data rate, large communication coverage, and other competing targets. It should be noted that it is almost impossible to meet all these targets simultaneously. Different application scenarios may focus on different specific targets, and there always exist some trade-offs among them. For instance, in vehicle-to-vehicle (V2V) communications [1], the major targets may include ultra-low latency, high reliability, and large communication coverage, while in machine-to-machine (M2M) communications, the major targets probably contain very low power consumption, high reliability, and large communication coverage for various wireless sensor networks (WSNs) and Internet of Things (IoT) applications [2]. Both the examples commonly have the two following targets, i.e., high reliability and large communication coverage, which can be achieved via cooperative relaying schemes [3]–[12].

The fundamental idea on cooperative relaying is that each relay forwards a replica of its received and/or further processed message to the next hop. The choices of the relay are, for instance, a base station, a device, a mobile terminal, a machine, or a vehicle depending on the application scenarios. The introduction of relays enables multiple routes, which connect the source to the destination, providing spatial diversity and thus enhancing the transmission reliability [8]. Since the data originated from the source is forwarded with the aid of the relays to the destination, obviously the communication distance can be increased compared to the case without any assistance of relays. However, the cooperative use of relays consumes more time slots for one entire transmission cycle, which degrades the network throughput. There exist various techniques which allow trade-offs, for example, among transmission reliability, network throughput, and communication coverage, determined by the number of engaged relays [13].
According to the operations at the relay, the relaying techniques can be mainly categorized into four major groups [7], [8], i.e., amplify-and-forward (AF) [14]–[16], compress-and-forward (CF) [17]–[19], compute-and-forward [10], [20], and decode-and-forward (DF) [7], [8], [21].

- **Amplify-and-Forward**: The fundamental idea of AF is that the relay amplifies the received signal and forwards it to the next hop. The drawback of AF is that the signal and noise are amplified simultaneously, and the noise components are accumulated, hop by hop. Thus, it does not achieve good performance, especially in the low signal-to-noise ratio (SNR) regime [7]. However, the implementation of AF is quite simple, which makes it attractive with a wide range of practical applications.

- **Compress-and-Forward**: In the CF relaying, the relay first compresses the received signal and then sends it to the next hop. The concept of compression may include, for instance, quantization and encoding with a rate-distortion code [22]. The CF strategy can achieve the capacity of some specific networks, e.g., dual-hop one-way relay and diamond networks [7], [23].

- **Compute-and-Forward**: Compute-and-Forward is tailored for multi-source relay network, aiming at harnessing the inter-user interference via structured codes, e.g., nested lattice codes [10], [20]. The core idea of compute-and-forward is that relays solve noise-corrupted linear functions of transmitted messages rather than treating the interference as noise, and then forward them to the destination. With sufficiently enough linear combinations, the destination is capable of obtaining its desired message [10], [20].

- **Decode-and-Forward**: The DF strategy utilizes another type of operations at the relay, compared to AF, CF, or compute-and-forward, where the received signal is first decoded, and then interleaved, re-encoded, and transmitted to the next hop if the decoded sequence is error free. The DF relaying strategy has been standardized in the Third Generation Partnership Project (3GPP) due to its robustness and implementation simplicity [24]. In addition, the DF strategy achieves the capacity of physically degraded relay channels [3]. However, in the DF relaying, the decoded sequence is discarded at the relay if decoding errors are detected. This results in a waste of resources, since the discarded sequence contains some relevant information about the source sequence. If it is transmitted to the destination, even better performance can be achieved because it can be utilized as a helper at the destination [25]. Moreover, for the sake of checking the recovery correctness at the relay, error detection, e.g., cyclic redundancy check (CRC), must be operated for the DF relaying. The use of CRC reduces the signaling spectrum efficiency.

A simple derivative technology from DF is selective DF (SDF) [26], where the relay decides either to forward the received sequence or to keep silence based on a predetermined criterion. Under this condition, SDF can achieve full diversity by forwarding the error-free sequences only, and avoid error propagation by not forwarding the erroneous sequences [26]. The original DF described in [5, Theorem 1] falls into the category of SDF where the criterion is based on CRC [26]. Other criteria by which the relay makes decisions are, for example, received SNR, and log-likelihood ratio (LLR), referred to as SNR-based SDF [27], [28] and LLR-based SDF [29], [30], respectively.

A new relaying strategy, referred to as lossy DF or simply, lossy forwarding (LF), has been developed from DF in [31] to address these two shortcomings of DF. Unlike the DF relaying, the received sequence at the relay nodes is decoded, re-encoded, and always forwarded to the next hop in LF relaying regardless of whether the decoding is successful or not. This is because the objective of LF is not detailed link design, but to achieve diversity gain and coverage extension. Better performance in terms of the outage probability, $\epsilon$-outage achievable rate, frame error rate (FER), and communication range can be achieved by LF compared to its DF counterpart. However, the disadvantages of LF are that the computational complexity due to the joint decoding (JD) at the destination is higher than that for DF relaying and LF is not supported by the existing standards. The destination in DF usually performs maximum ratio combining (MRC) of the received signals from the relays which successfully recover the information sequence from the previous hops. Better FER performance can be achieved if the destination implements JD other than MRC in the DF relaying [32]. Under the fair assumption that the destination conducts JD for both LF and DF relaying, the computational complexity for LF is slightly higher than that for DF at the destination. However, the shortcoming with LF can be compensated for by the various advantages, especially the performance improvements. In certain practical use cases where all the wireless channels suffer from a deep fade, LF relaying can address the issue of always-be-silent of the relays and allow the destination to obtain a “lossy” or even “lossless” copy of the original source sequence. This can be particularly beneficial for delay-critical applications. Obviously, LF is less sensitive to hardware impairment applications. Obviously, LF is less sensitive to hardware impairment compared to AF [33], because LF does not count on the coherence in the analog domain.

The theoretical performance bounds of the three-node LF relaying network were calculated, for example, in [34]–[37] based on the combination of the Slepian-Wolf theorem [38] and the Shannon’s lossy source-channel separation theorem [22]. However, in [39]–[41], it was realized that the achievable rate region can be expanded by eliminating the necessity of recovering the sequence sent from the relay. It was further found in [25], [42] that the theoretical analysis of the LF relaying system can utilize the theorem of source coding with a helper [43]–[45], because only the original information from the source needs to be recovered at the destination and the sequence from the relay is purely considered as a helper. This opened a new research field for which a lot of academic work has been done and important results have been published. The outage probability analyses for the three-node LF relaying network were conducted in [25], [39]–[42].
Multiple access relay channel (MARC) [64]–[67] is another form of cooperative wireless network, where the relay performs network coding, e.g., bit-wise exclusive or (XOR), on the information sequences from multiple sources, and forwards the XORed copy to the destination. The work in [64]–[67] was based on DF relaying. In this regard, if the relay fails to recover all the information sequences sent from the sources, it will discard all the decoded sequences rather than perform network coding on them, leading to a waste of resources. In order to avoid it, LF in orthogonal MARC was intensively studied in [68]–[72] in terms of achievable rate region, outage probability, and JD, with its extension to non-orthogonal MARCs in [69], [73].

The LF principle can also be applied to more complicated relaying networks, e.g., two-way relay network [74] and general multi-source multi-relay network [75], [76].

The historical footprint of the major contributions under the LF relaying framework are summarized in Table I, listed in chronological order. The major focus points of this tutorial paper are on representing characteristics of LF relaying, including achievable rate regions, theoretical performance limits for various network setups, performance evaluation by exploiting practical coding-decoding chain, and comparison with DF relaying scheme. Nevertheless, we also introduce some related research topics, for instance, relay selection for the scenario in the presence of multiple relays, upper-layer issues for the one-way relay network, and full duplex in two-way relay network. The network setups include the following five basic exemplifying scenarios:

(a) Three-node half-duplex relay network;
(b) Single-source multi-relay network with direct S-D link;
(c) MARC;
(d) Two-way relay network;
(e) General multi-source multi-relay network.

In the paper, we provide a comprehensive survey of the state-of-the-art on LF cooperative relaying. More specifically, we start by introducing some information-theoretic preliminaries, e.g., distributed lossless source coding and distributed lossless source coding with a helper, in Section II. Practical low-complexity coding-decoding chain for LF relay networks is described in Section III, where for the channel code we utilize a very simple accumulator (ACC) aided turbo code from the class of serially concatenated convolutional codes (SCCCs). This coding-decoding structure is used to evaluate the theoretical results obtained via numerical calculations. Section IV analyzes the network models (a) and (b). Network models (c), (d), and (e) are the investigation topics in Sections VI, V, and VII, respectively. Section VIII summarizes the paper and sheds lights on some future research directions.

The list of acronyms used in the paper is shown in Table II.

II. INFORMATION-THEORETIC PRELIMINARIES

In this section, we present some critical information-theoretic preliminaries, which will be utilized in the achievable rate region and performance limit analyses of the LF relaying networks. To be specific, we provide the achievable rate regions for distributed lossless source coding and distributed

---

1 In the single-source multi-relay networks, distributed space-time coding [54] is another technique to achieve full diversity of such networks. It should be noted that the combination of distributed space-time coding and LF can further bring coding gain.
lossless source coding with a helper. Also, we introduce the Shannon’s lossy source-channel separation theorem, which is used to measure the lower bound of distortion in the source-to-relay (S-R) as well as relay-to-destination (R-D) link transmissions.

A. Distributed Lossless Source Coding

In this subsection, we introduce two important theoretical frameworks related to distributed lossless source coding. The first one considers the transmission of multiple correlated sources, where all the sources need to be recovered at the destination with arbitrarily small error probability. The second one considers the transmission of multiple correlated sources with a helper. The difference in the two cases lies in the fact that the helper does not need to be recovered in the second case; its role is just to assist the recovery of the correlated sources.

1) Slepian-Wolf Coding: Distributed lossless source coding system is depicted in Fig. 1. Let \( \{X_1^i, X_2^i, \ldots, X_N^i\}_{i=1}^{\infty} \) be a sequence of jointly independent and identically distributed (i.i.d.) discrete random tuples generated by the sources \( X_1, X_2, \ldots, X_N \) with a joint probability distribution function \( p(x_1, x_2, \ldots, x_N) \). According to the source coding theorem [86], if all the sources are encoded together, a sum rate \( H(X_1, X_2, \ldots, X_N) \) is sufficient to achieve lossless decoding, where \( H(X_1, X_2, \ldots, X_N) \) is the joint entropy of \( X_1, X_2, \ldots, X_N \). However, if the sources are independently encoded with rates \( R_1, R_2, \ldots, R_N \), we get the following achievable rate region originally introduced in [38]:

\[
\sum_{j \in S} R_j \geq H(X(S)|X(S^c)), \quad \forall S \subseteq \{1, 2, \ldots, N\},
\]

where \( H(A|B) \) denotes the entropy of \( A \) conditioned on \( B \).

2) Distributed Lossless Source Coding with a Helper: The Slepian-Wolf coding assumes that all the correlated sources need to be recovered at the destination. However, in coopera-

---

**TABLE I: SUMMARY OF MAJOR CONTRIBUTIONS ON LF RELAYING**

<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>Garcia-Frias and Zhao [35]</td>
<td>Considered JSSC of two correlated binary information sequences with correlation information estimation at the destination.</td>
</tr>
<tr>
<td>2005</td>
<td>Hu and Li [77]</td>
<td>Proposed Slepian-Wolf cooperation, which exploits distributed source coding technologies in wireless cooperative communications.</td>
</tr>
<tr>
<td>2006</td>
<td>Li et al. [78]</td>
<td>Proposed a distributed turbo coding scheme with soft information relaying. Instead of making decision on the transmitted information sequence at the relay, the relay calculates and forwards the soft estimates of the transmitted sequence.</td>
</tr>
<tr>
<td>2007</td>
<td>Woldegebreal and Karl [79]</td>
<td>Considered a network coding based MARC in the presence of non-ideal source-to-relay links, and analyzed the outage performance and communication coverage.</td>
</tr>
<tr>
<td>2007</td>
<td>Sneessens et al. [80]</td>
<td>Developed a decoding algorithm which takes into account the error probability incurred in the source-to-relay link.</td>
</tr>
<tr>
<td>2012</td>
<td>Anwar and Matsumoto [31]</td>
<td>Proposed an iterative decoding technique, called accumulator (ACC) aided turbo code, where the source-relay correlation is estimated and exploited during the JD at the destination.</td>
</tr>
<tr>
<td>2014</td>
<td>Zhou et al. [25]</td>
<td>Derived the exact outage probabilities for the three-node network by utilizing the theorems of lossy source-channel separation and source coding with a helper.</td>
</tr>
<tr>
<td>2015</td>
<td>Wolf et al. [55]</td>
<td>Proposed a suboptimal power allocation strategy for a two-relay system using convex optimization to minimize the outage probability.</td>
</tr>
<tr>
<td>2016</td>
<td>Jayakody et al. [81]</td>
<td>Proposed a framework for employing spatially-coupled low-density parity-check (LDPC) codes with the threshold-based LF protocol over AWGN and Rayleigh fading channels in MARCs.</td>
</tr>
<tr>
<td>2016</td>
<td>Brulatout et al. [82]</td>
<td>Proposed a medium access control layer protocol for LF relaying and introduced testbed implementation based on universal software radio peripheral (USRP) and GNU radio frameworks.</td>
</tr>
<tr>
<td>2016</td>
<td>Kosek-Szot et al. [83]</td>
<td>Proposed a centralized medium access protocol specifically designed to operate with LF relaying for coordinated wireless local area networks (WLANs).</td>
</tr>
<tr>
<td>2017</td>
<td>González et al. [84]</td>
<td>Designed a simple, yet efficient power allocation policy for a distributed source coding/joint-decoding scheme operating over a DF relaying system which allows intra-link errors.</td>
</tr>
<tr>
<td>2017</td>
<td>He et al. [73]</td>
<td>Derived the outage probability for non-orthogonal MARCs with LF relaying, and simulated the FER performance using ACC aided turbo codes.</td>
</tr>
<tr>
<td>2018</td>
<td>González et al. [85]</td>
<td>Designed a new power allocation policy that is applicable to any arbitrary fading scenarios.</td>
</tr>
</tbody>
</table>

---

**Fig. 1:** Distributed lossless source coding of an arbitrary number of sources.

**Theorem 1 (Slepian-Wolf Theorem):** Let \( H(X(S)) \) denote the joint entropy of the tuple \( (X_j : j \in S) \) and \( S^c \) denote the complement of the set \( S \). A sequence \( \{X_1^i, X_2^i, \ldots, X_N^i\}_{i=1}^{\infty} \) of discrete random tuples is drawn according to the joint probability distribution function \( p(x_1, x_2, \ldots, x_N) \), where \( X_n^i \in X_n \), \( n = 1, 2, \ldots, N \). Then, the achievable rate region for the lossless recovery of \( (X_1, X_2, \ldots, X_N) \) is expressed as

\[
\sum_{j \in S} R_j \geq H(X(S)|X(S^c)), \quad \forall S \subseteq \{1, 2, \ldots, N\},
\]
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>accumulator</td>
</tr>
<tr>
<td>AF</td>
<td>amplify-and-forward</td>
</tr>
<tr>
<td>AN</td>
<td>artificial noise</td>
</tr>
<tr>
<td>AODV</td>
<td>ad hoc on-demand distance vector</td>
</tr>
<tr>
<td>AWGN</td>
<td>additive white Gaussian noise</td>
</tr>
<tr>
<td>BCJR</td>
<td>Bahl-Cocke-Jelinek-Raviv</td>
</tr>
<tr>
<td>BEP</td>
<td>bit error probability</td>
</tr>
<tr>
<td>BER</td>
<td>bit error rate</td>
</tr>
<tr>
<td>BICM-ID</td>
<td>bit interleaved coded modulation with iterative detection</td>
</tr>
<tr>
<td>BPSK</td>
<td>binary phase-shift keying</td>
</tr>
<tr>
<td>CCC</td>
<td>constellation constrained capacity</td>
</tr>
<tr>
<td>CEO</td>
<td>chief executive officer</td>
</tr>
<tr>
<td>CFP</td>
<td>contention free period</td>
</tr>
<tr>
<td>CLF-MAC</td>
<td>centralized lossy forwarding-medium access control</td>
</tr>
<tr>
<td>CP</td>
<td>contention period</td>
</tr>
<tr>
<td>CRC</td>
<td>cyclic redundancy check</td>
</tr>
<tr>
<td>CSI</td>
<td>channel state information</td>
</tr>
<tr>
<td>CSMA/CA</td>
<td>carrier sense multiple access with collision avoidance</td>
</tr>
<tr>
<td>D2D</td>
<td>device-to-device</td>
</tr>
<tr>
<td>DF</td>
<td>decode-and-forward</td>
</tr>
<tr>
<td>DSR</td>
<td>dynamic source routing</td>
</tr>
<tr>
<td>DTC</td>
<td>distributed turbo code</td>
</tr>
<tr>
<td>EXIT</td>
<td>extrinsic information transfer</td>
</tr>
<tr>
<td>FER</td>
<td>frame error rate</td>
</tr>
<tr>
<td>GCC</td>
<td>Gaussian codebook capacity</td>
</tr>
<tr>
<td>GI</td>
<td>global iteration</td>
</tr>
<tr>
<td>HBC</td>
<td>hybrid broadcast</td>
</tr>
<tr>
<td>i.i.d.</td>
<td>independent and identically distributed</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>JD</td>
<td>joint decoding</td>
</tr>
<tr>
<td>JSCC</td>
<td>joint source-channel coding</td>
</tr>
<tr>
<td>LF</td>
<td>lossy forwarding</td>
</tr>
<tr>
<td>LI</td>
<td>local iteration</td>
</tr>
<tr>
<td>LLR</td>
<td>log-likelihood ratio</td>
</tr>
<tr>
<td>MM</td>
<td>machine-to-machine</td>
</tr>
<tr>
<td>MARC</td>
<td>multiple access relay broadcast</td>
</tr>
<tr>
<td>MAC</td>
<td>multiple access channel</td>
</tr>
<tr>
<td>MARC</td>
<td>multiple access relay channel</td>
</tr>
<tr>
<td>MEC</td>
<td>mobile edge computing</td>
</tr>
<tr>
<td>MEU</td>
<td>mobile end user</td>
</tr>
<tr>
<td>MIMO</td>
<td>multiple-input multiple-output</td>
</tr>
<tr>
<td>MRC</td>
<td>maximum ratio combining</td>
</tr>
<tr>
<td>MRS</td>
<td>multiple-relay selection</td>
</tr>
<tr>
<td>MRT</td>
<td>maximum ratio transmission</td>
</tr>
<tr>
<td>MUD</td>
<td>multilayer detection</td>
</tr>
<tr>
<td>OLSR</td>
<td>optimized link state routing</td>
</tr>
<tr>
<td>OSTBC</td>
<td>orthogonal space-time block code</td>
</tr>
<tr>
<td>pdf</td>
<td>probability density function</td>
</tr>
<tr>
<td>QoS</td>
<td>quality of service</td>
</tr>
<tr>
<td>RR</td>
<td>resource reservation</td>
</tr>
<tr>
<td>SCCC</td>
<td>serially concatenated convolution code</td>
</tr>
<tr>
<td>SDJ</td>
<td>selective decode-and-forward</td>
</tr>
<tr>
<td>SDN</td>
<td>software defined networking</td>
</tr>
<tr>
<td>SER</td>
<td>symbol error rate</td>
</tr>
<tr>
<td>SIC</td>
<td>successive interference cancellation</td>
</tr>
<tr>
<td>SISO</td>
<td>soft-input soft-output</td>
</tr>
<tr>
<td>SNR</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>SRCC</td>
<td>systematic recursive convolution code</td>
</tr>
<tr>
<td>SRS</td>
<td>single-relay selection</td>
</tr>
<tr>
<td>TAS</td>
<td>transmit antenna selection</td>
</tr>
<tr>
<td>TDBC</td>
<td>time division broadcast</td>
</tr>
<tr>
<td>TDMA</td>
<td>time division multiple access</td>
</tr>
<tr>
<td>USRP</td>
<td>universal software radio peripheral</td>
</tr>
<tr>
<td>V2V</td>
<td>vehicle-to-vehicle</td>
</tr>
<tr>
<td>WLAN</td>
<td>wireless local area network</td>
</tr>
<tr>
<td>WSN</td>
<td>wireless sensor network</td>
</tr>
<tr>
<td>XOR</td>
<td>exclusive-OR</td>
</tr>
</tbody>
</table>

**TABLE II: LIST OF ACRONYMS USED IN THIS PAPER**

**Fig. 2:** Distributed lossless source coding of an arbitrary number of sources with a helper.

Theorem 2 (Distributed Lossless Source Coding with a Helper): Let \( \{X_1, X_2, \ldots, X_N, Y\} \) be a sequence of jointly i.i.d. discrete random tuples generated by the sources \( X_1, X_2, \ldots, X_N, Y \). The achievable rate region for the lossless recovery of \( (X_1, X_2, \ldots, X_N) \) with the assistant of \( Y \) is given by

\[
\sum_{j \in S} R_j \geq H(X(S)|U, X(S')) \quad \forall S \subseteq \{1, 2, \ldots, N\}, \quad R_{N+1} \geq I(Y; U),
\]

(2)

for some conditional probability mass function (pmf) \( p(u|y) \) with \(|U| \leq |Y| + 2^N - 1 \). Here, \( I(A; B) \) denotes the mutual information between \( A \) and \( B \).

The proof of cardinality bound of \( \mathcal{U} \) can be obtained using the convex cover method in [87].

Source coding with a helper is a special case of distributed lossless source coding with a helper when \( N = 1 \).

**B. Shannon’s Lossy Source-Channel Separation Theorem**

The concept of LF allows intra-link errors (i.e., decoding errors occur in the S-R links) in the network [25], [81]. The information transmission in the S-R links can therefore be modeled by the Shannon’s lossy source-channel separation theorem [22]. The theorem states that the transmission of a message from a transmitter to a receiver with a distortion \( D \), needs to satisfy

\[
R_c R(D) \leq C(\gamma),
\]

(3)

where \( R_c \) is the normalized spectrum efficiency taking the channel coding rate and modulation order into consideration, and even source compression rate, \( R(D) \) is the rate-distortion function, and \( C(\gamma) \) denotes the Gaussian channel capacity.
(strictly speaking, it should be the constellation constrained capacity (CCC), but the use of Gaussian capacity does not make any significant difference in the outage calculation [53]) as a function of received SNR $\gamma$. Similarly, the distortion occurred in the helper transmission (i.e., R-D link) can also be calculated by following the Shannon’s lossy source-channel separation theorem.

In the LF relaying, we focus on discrete binary sources having zero/one extrinsic probabilities of 0.5, the distortion is expressed as

$$R(D) = 1 - H_2(D),$$

where $H_2(D) = -D \log_2(D) - (1 - D) \log_2(1 - D)$ is the binary entropy function [88]. With the help of (3) and (4), we can derive the relationship between the distortion (under Hamming distortion measure in the paper) and received SNR, as

$$D_{\text{min}} = \begin{cases} H^{-1}_2(1 - C(\gamma)/R_c), & \text{for } 0 \leq \gamma < \gamma^* \\ 0, & \text{for } \gamma \geq \gamma^* \end{cases},$$

where $\gamma^*$ is the minimum SNR required for zero distortion and $H^{-1}_2(\cdot)$ is the inverse binary entropy function.

III. PRACTICAL CODING-DECODING ALGORITHMS

In this section, we introduce a practical, simple, and efficient coding-decoding scheme which can approach the theoretical performance bounds derived in the following sections. First, we introduce the encoder structure, which is in principle a SCCC. Then, we provide a JD scheme to be used in the multi-relay networks with LF relaying, which are the three-node relay network as a special case. Modifications are required when extending the JD to other relaying networks, for example, MARC, to be investigated in Section V.

A. Encoder Structure

The single-node encoding structure is presented in Fig. 3, which belongs to the class of SCCCs. It was first devised in [57] for the sake of eliminating error floor in bit interleaved coded modulation with iterative detection (BICM-ID) systems [89]. The encoding structure consists of five components, one rate-$R_c$ outer code, two interleavers, one rate-1 inner code (i.e., ACC), and one modulator (i.e., signal mapper). The overall rate of the channel encoder is $R_c$. The binary information sequence $u$ is first input to an interleaver $\Pi_1$ before processed by the outer encoder. This interleaver can be used to achieve independent extrinsic information for different sources when jointly decoding the transmitted sequences from the sources. Furthermore, the interleaver $\Pi_1$ can also be used for jointly decoding multiple retransmitted sequences from the same source as long as different interleaving patterns are used for each re-transmission [32].

1) Outer code: The utmost component, i.e., the outer code, is a memory one systematic non-recursive convolutional code (SNRCC). We use a simple code, because our focus is on the outage probability and diversity order, not on the code optimization. Nevertheless, we performed a series of EXIT analyses and found that a serial combination of a memory one SNRCC and ACC, i.e., ACC aided turbo code, provides very good EXIT curve matching and hence very sharp turbo cliff can be achieved in the AWGN channels. Moreover, in all the exemplifying scenarios, the proposed code was shown to be limit-approaching over Rayleigh block fading channels [72], [73], [90].

2) Inner code: After the encoding of the outer code, the output sequence is further interleaved and forwarded to the encoder of inner code. The inner code is a doped ACC [57], which is memory one NSRCC (equivalent to differential encoder if the output is all coded bits). Inner doping is implemented by replacing some of the uncoded bits by the coded bits for the purpose of enhancing the performance of the demodulator [57],[91]. The doping rate is usually expressed by a positive integer $P$, which means that every $P$th uncoded bit is removed from the ACC output and replaced by the coded bits. With this technique, the EXIT curve matching falls into a matter of doping rate selection and hence can be easily managed [57],[91]. The most beneficial point of the combination of SNRCC and ACC is that the decoding process of ACC can be replaced by a simple differential detector if $P = 1$ (full accumulation) and systematic information sequence can simply be extracted without performing any decoding process for SNRCC.

B. Encoding-Decoding of Multi-Relay Network

A schematic diagram of a multi-relay network is shown in Fig. 4, where $u_k = u \oplus e_k$ denotes the decoded sequence at the relay $k$, $k \in \{1,2,\cdots,K\}$, $u$ is the original binary i.i.d. information sequence from the source, $\oplus$ is modulo-2 addition, and $e_k = [e_{k,1}, e_{k,2}, \ldots, e_{k,L}]$ is a bit-flipping sequence defined as

$$e_{k,n} = \begin{cases} 1, & \text{with probability } p_k, \\ 0, & \text{with probability } 1 - p_k, \quad n = 1,2,\ldots,L \end{cases},$$

where $p_k$ is determined by the channel condition between the source and relay $k$ and can be computed by following (3). $L$ is the length of the information sequence. An alternative interpretation of (6) is that $u$ and $u_k$ are input and output of a binary symmetric channel (BSC) with crossover probability $p_k$. Obviously, the smaller the $p_k$ value, the higher the correlation coefficient $p_k$ between $u$ and $u_k$ with $p_k = 1 - 2p_k$ [31].

In the multi-relay network, we assume that a direct link exists between the source and the destination. Under such an assumption, $u$ is directly input to ENC0 without adding any bit-flipping sequence, as shown in Fig. 4. The decoded

\[ u \longrightarrow \Pi_1 \longrightarrow C \longrightarrow \Pi_2 \longrightarrow \text{ACC} \longrightarrow \text{Modulation} \]

Fig. 3: Single-node encoder structure.
sequence $u_k$ at the relay $k$ is interleaved and encoded by the encoder presented in Fig. 3. The coded and modulated sequence $s_k$ is corrupted by AWGN noise $n_k$.

A joint decoder of a multi-relay system is presented in Fig. 5, consisting of local iterations (LI) and global iteration (GI). Iterative soft-input soft-output (SISO) decoding, well known already from the framework of turbo detection/decoding techniques, including BICM-ID, is performed for each parallel branch. The SISO components are depicted by the white boxes in Fig. 5, i.e., the extrinsic information in the form of LLR exchange between the combined block of demapper and ACC decoder (denoted as ACC\(^{-1}\) in Fig. 5), and the decoder of SNRCC (e.g., $D_k$ in Fig. 5). As stated in Section III-A2, ACC is a rate-1 convolutional code, thus the BCJR algorithm [92] needs to be performed. Because of the fact that the information sequences transmitted from the source and relays are correlated, the correlation information\(^3\) is utilized to compensate for the LLR value during each GI. Further information on estimating the correlation information at the destination can be found in [31], [35].

The decoding process is summarized below:

1) Demapping: calculate the LLR of the demapper output using SISO demapping presented in [93], [94].

2) Perform LI between ACC\(^{-1}\) and $D_k$, $k = 0, \ldots, K$.

3) The extrinsic LLRs $L_{e}(u_k)$, which are obtained by subtracting the a priori LLRs $L_{a}(u_k)$ from the a posteriori LLRs $L_p(u_k)$, are deinterleaved by $\Pi_{k,1}$ and then updated through the $f_c(\cdot)$ as in [31], [71], which modifies the systematic extrinsic LLR by taking into account the knowledge of the correlation information $p_k$, as

$$L_{e}(u, k) = f_c(\Pi_{k,1}^{-1}(L_{e}(u))), p_k).$$  (7)

4) After enough rounds of LI, GI is invoked where a degree $K + 1$ variable node combines $K$ extrinsic LLR vectors $L_a(u, j), j \neq k$, to calculate the message to be fed back to the decoder $D_k$. Thus, the a priori LLR $L_a(u, k)$ is then obtained as

$$L_a(u, k) = \sum_{j \neq k} L_e(u, k),$$  (8)

where $\kappa = \{0, 1, 2, \ldots, K\}$ is the set of encoder indices, and $k' \kappa$ indicates removal of $k$ from the set $\kappa$.

5) The output of $f_c(L_a(u, k), p_k)$ is interleaved by $\Pi_{k,1}$ to form the a priori LLR $L_a(u_k)$, which is directly fed back to $D_k$.

The entire JD process is conducted in an iterative manner where Steps 2) to 5) are repeated with the switches in Fig. 5 closed except in the final iteration. In the final GI, the switches are open to combine all the information to produce the a posteriori LLR. A hard decision is made on $\hat{u}$ to obtain a final estimate of $u$, i.e., $\hat{u}$, based on the a posteriori LLR $L_p(u)$.

The coding-decoding chain described above can be applied to the exemplifying scenarios (a) and (b). For the scenarios of MARC, two-way relay network, and general multi-source multi-relay network, the encoding chain is the same as that shown in Fig. 3. However, due to the application of network coding at the relays, modifications were made to the JD at the destination, for instance, in MARC [72], [73]. Detailed information will be provided in Section V on JD in MARC. The key roles played by the JD are the LLR modification by $f_c(\cdot)$ and LLR information exchange among different decoders. As long as we can derive the $f_c(\cdot)$ function corresponding to the network, and formulate the extrinsic LLR information, the extension of JD to general multi-source multi-relay networks is straightforward. Obviously, the BCJR algorithm [92] has to be performed in the LI to calculate the LLRs.

### IV. ONE-WAY RELAY NETWORK

In this section, we mainly concentrate on the theoretical performance limits and upper-layer issues of the three-node LF relaying network, depicted in Fig. 6, and briefly introduce single-source multi-relay network with a focus on relay selection. With the introduction of LF relaying protocol, the relay always forwards its decoded sequence to the destination despite of the decoding errors.

The pioneering work on the performance limit analyses of the three-node LF relaying system was provided by Cheng et al. [41], where the S-R link is represented by a BSC with time-invariant crossover probability (also known as bit-flipping probability) and the remaining links suffer from i.i.d. Rayleigh block fading. Zhou et al. [25] further relaxed the constraint on the S-R link and assumed that it also suffers from i.i.d. Rayleigh block fading. Extension to Nakagami-$m$ fading and temporally and spatially correlated fading was investigated in [52], [53]. The brief summary of the related work on the one-way relay network can be found in Table III.

In [41], the achievable rate region was first obtained by following the Slepian-Wolf theorem. The achievable rate region was further expanded in [41], due to the fact that the sequence transmitted from the relay does not need to be recovered at the destination. Finally, in [25], the authors...
found that the problem exactly fits for the source coding with a helper. The results indicate that the achievable rate region in [41] is an approximation of that in [25] and the approximation is very tight. The achievable rate region drawn by following the theorem of source coding with a helper is the largest among the three. Accordingly, the smallest outage probability is expected. The evolution towards more accurate interpretation of achievable rate region is presented in Fig. 7.

A. Channel Type

As shown in Table III, different types of channel are taken into consideration when investigating the performance limits of the one-way relay network with LF at the relay. BSC was utilized to characterize the S-R link and research was conducted to analyze the influence of source-relay correlation on the outage probability and diversity order [41]. Rayleigh block fading was assumed in many publications, e.g., [40], [41] for S-D and R-D links, and [25], [52] for all the links. Extension to Nakagami-m block fading was studied in [53].

- BSC: The BSC builds the relationship between the binary sequences sent from the source and the decoded sequence from the relay [41]. The crossover probability $p$ represents the strength of the source-relay correlation. Two extreme cases are: 1) $p = 0$ stands for full correlation,
Achievable rate region

Fig. 7: Achievable rate regions: (a) Slepian-Wolf rate region [41], (b) Approximated rate region [25], [40], [41], and (c) Exact rate region derived from the theorem for source coding with a helper [25], [52], [53].

the binary sequence from the source and the decoded sequence from the relay are exactly the same; 2) \( p = 0.5 \) stands for no correlation, i.e., the binary sequence from the source and the decoded sequence from the relay are mutually independent.

- **Rayleigh Block Fading:** For the Rayleigh channel, the absolute value of the channel coefficient is Rayleigh distributed. Rayleigh block fading has been widely used in the literatures. The probability density function (pdf) of received SNR \( \gamma \) over the Rayleigh fading channel follows [95]

\[
p(\gamma) = \frac{1}{\bar{\gamma}} \exp\left(-\frac{\gamma}{\bar{\gamma}}\right),
\]

where \( \bar{\gamma} \) is the average received SNR for the Rayleigh fading channel.

- **Nakagami-\( m \) Block Fading:** The pdf of the received SNR over the Nakagami-\( m \) fading channel follows

\[
p(\gamma) = \frac{m^m \gamma^{m-1}}{\bar{\gamma}^m \Gamma(m)} \exp\left(-\frac{m \gamma}{\bar{\gamma}}\right), \quad m > 0.5,
\]

where \( \Gamma(\cdot) \) is the Gamma function, and shape factor \( m \) determines the severity of the Nakagami-\( m \) channel. When \( m = 1 \), the Nakagami-\( m \) fading channel degrades into Rayleigh fading channel. The larger of the value \( m \), the less severe of the fading variation.

With the block fading assumption, e.g., Rayleigh block fading and Nakagami-\( m \) block fading, we can calculate the distortion, i.e., bit error probability (BEP), for the binary information sequence within each block by exploiting (5). The BEP keeps fixed within one block but changes from one block to another. The S-R link can be formulated by a series of BSCs with time varying crossover probabilities (equivalent to BEPs), which are determined by the block channel condition.

**B. Achievable Rate Region Analysis**

Assuming that the S-D and R-D links can support the rates \( R_1 \) and \( R_2 \), respectively, the achievable rate regions considered in [25], [40], [41], [52], [53] are shown in Fig. 7, where \( u_1 \) denotes the binary uniformly distributed information sequence from the source, \( u_2 \) denotes the decoded binary information sequence at the relay, and the correlation information between \( u_1 \) and \( u_2 \) is \( p \).

In the initial work [41], the achievable rate region was first determined by the Slepian-Wolf rate region. Since the message from R does not need to be recovered and it purely serves as a helper, the achievable rate region is extended by adding the region determined by \( R_1 \geq H(u_1) \) and \( 0 \leq R_2 \leq H(u_2|u_1) \) [25], [40], [41], as shown in Fig. 7 (b). However, the exact achievable rate region is obtained based on the theorem of source coding with a helper, which was found in [25]. Comparison was made between the cases of Fig. 7 (b) and Fig. 7 (c) in [25], which concludes that they achieve almost the same performance in terms of outage probabilities. In the latest publications [52], [53] on the three-node LF relaying system, the authors rely on the theorem of source coding with a helper.
In the following, we mathematically derive the exact and approximated achievable rate regions, as shown in Fig. 7 (c) and (b), respectively.

1) Exact Rate Region: In the three-node LF relaying system, the recovery of $\mathbf{u}_1$ sent from the source is the main objective of the destination. In other words, $\mathbf{u}_1$ sent from the relay does not have to be perfectly decoded at the destination. The coding-decoding of $\mathbf{u}_1$ and $\mathbf{u}_2$ can be formulated as a source coding with a helper problem [87], where two correlated sources are separately encoded but jointly decoded at the destination, and only one of them needs to be recovered and the other acts as a helper. Apparently, in the three-node LF relaying system, $\mathbf{u}_2$ helps to decode $\mathbf{u}_1$ at the destination, as shown in Fig. 8 (a). Following Theorem 2, error-free decoding of $\mathbf{u}_1$ at the destination with the aid of $\mathbf{u}_2$ can be realized if the rate pair $(R_1, R_2)$ satisfies

$$\begin{cases} R_1 & \geq H(\mathbf{u}_1; \hat{\mathbf{u}}_2), \\ R_2 & \geq I(\mathbf{u}_2; \hat{\mathbf{u}}_2), \end{cases}$$

(11)

with $\hat{\mathbf{u}}_2$ being the decoded copy of $\mathbf{u}_2$ at the destination. Similarly, the relationship between $\mathbf{u}_2$ and $\hat{\mathbf{u}}_2$ can also be characterized by the bit-flipping model (or BSC) with bit-flipping probability (or crossover probability) $\alpha$, where $\alpha \in [0, 0.5]$. It is not difficult to figure out that $H(\mathbf{u}_1; \hat{\mathbf{u}}_2) = H_2(\alpha + \beta)$ and $I(\mathbf{u}_2; \hat{\mathbf{u}}_2) = H(\hat{\mathbf{u}}_2) - H(\hat{\mathbf{u}}_2|\mathbf{u}_2) = 1 - H_2(\alpha)$, with $\alpha + \beta = (1 - \alpha)p + \alpha(1 - p)$ being the binary convolution of $\alpha$ and $p$ [25].

When $\alpha = 0$, the conditions in (11) become $R_1 \geq H_2(0|p) = H_2(p)$ and $R_2 \geq 1 - H_2(0) = 1$. When $\alpha = 0.5$, the conditions in (11) become $R_1 \geq H_2(0.5|p) = 1$ and $R_2 \geq 1 - H_2(0.5) = 0.5$. When $0 < \alpha < 0.5$, the conditions in (11) become $R_1 \geq H_2(\alpha + \beta)$ and $R_2 \geq 1 - H_2(\alpha)$. Based on the description above, we can further write (11) as

$$R_1 \geq \begin{cases} H_2(\beta), & \text{for } R_2 \geq 1, \\ H_2[H_2^{-1}(1 - R_2) + \alpha|p], & \text{for } 0 \leq R_2 \leq 1. \end{cases}$$

(12)

2) Approximated Rate Region: As shown in (12), there appears a non-linear border determined by $R_1 \geq H_2[H_2^{-1}(1 - R_2) + \alpha|p]$, for $0 \leq R_2 \leq 1$. In this subsection, we introduce an approximated rate region, for which all the borderlines are constrained by linear functions. The approximated rate region is obtained based on the Slepian-Wolf theorem. Different from the theorem of source coding with a helper, all the correlated sources need to be recovered losslessly according to the Slepian-Wolf theorem. The achievable rate region of the three-node LF relaying system can be approximated as the Slepian-Wolf region plus an additional region, and the approximation was proven to be quite tight in [25].

For better illustration of the approximated rate region, we first consider the lossless transmission of both $\mathbf{u}_1$ and $\mathbf{u}_2$ from the perspective of the Slepian-Wolf theorem, abstracted in Fig. 8 (b). Following Theorem 1, lossless recovery of both $\mathbf{u}_1$ and $\mathbf{u}_2$ at the destination with JD can be realized if the rate pair $(R_1, R_2)$ satisfies

$$\begin{cases} R_1 & \geq H(\mathbf{u}_1; \mathbf{u}_2), \\ R_2 & \geq H(\mathbf{u}_2; \mathbf{u}_1), \\ R_1 + R_2 & \geq H(\mathbf{u}_1, \mathbf{u}_2). \end{cases}$$

(13)

Since the source $\mathbf{u}_1$ is uniformly distributed binary sequence, it is readily to obtain that $H(\mathbf{u}_1) = H(\mathbf{u}_2) = 1$, $H(\mathbf{u}_1; \mathbf{u}_2) = H(\mathbf{u}_2|\mathbf{u}_1) = H_2(p)$, and $H(\mathbf{u}_1, \mathbf{u}_2) = 1 + H_2(p)$. The approximated rate region takes into consideration of the additional region described by $R_1 \geq H(\mathbf{u}_1) = 1$ and $0 \leq R_2 \leq H(\mathbf{u}_2|\mathbf{u}_1) = H_2(p)$, along with the region defined by (13). It can be further expressed as

$$R_1 \geq \begin{cases} H_2(p), & \text{for } R_2 \geq 1, \\ 1 + H_2(p) - R_2, & \text{for } 0 \leq R_2 \leq H_2(p), \end{cases}$$

(14)

where all the borderlines are determined by linear functions.

It is rather straightforward to prove that the achievable rate region by following the source coding with a helper is larger than that by following the Slepian-Wolf theorem. Therefore, the outage probability calculated by following source coding with a helper is smaller than that calculated by following the Slepian-Wolf theorem. These statements have been verified in [25].

C. Outage Probability Analysis

1) Relationship Between $R_1$, $R_2$ and Their Corresponding Instantaneous Received SNRs: The Shannon’s source-channel separation theorem builds the relationship among $R_1$, $R_{c,1}$, and $\gamma_1$, and also that among $R_2$, $R_{c,2}$, and $\gamma_2$. If the following conditions hold [35]

$$\begin{cases} R_1 & \leq C(\gamma_1), \\ R_2 & \leq C(\gamma_2), \end{cases}$$

(15)

the error probability of the S-D and R-D links can be made arbitrarily small. Here, $R_{c,1}$ and $R_{c,2}$ indicate the normalized spectrum efficiencies of S-D and R-D links, respectively, $\gamma_1$ and $\gamma_2$ are the instantaneous received SNR for the S-D and R-D links, respectively.

In the theoretical outage probability analysis, we consider the equality of (15). Therefore, we can write the rate $R_i$ as a function of the corresponding instantaneous received SNR $\gamma_i$ as

$$R_i = \frac{C(\gamma_i)}{R_{c,i}}, \text{ for } i = 1, 2.$$ 

(16)

2) Outage Probability Calculation: With the assumption of block fading, the channel variation keeps fixed within one transmission block. That is to say, the value $p$ keeps fixed, and achievable rate region also keeps fixed. The calculation of $p$ is described in Section II-B. According to (16), the rate pair for recovering the source with arbitrarily small error probability is computed. If the calculated rate pair falls outside of the achievable rate regions, either the exact rate region or the approximated rate region, an outage event happens. Based on the obtained exact and approximated achievable rate regions in Section IV-B, the exact and approximated outage probability can be expressed by

$$P_{out}^{\text{Exact}} = \Pr \{0 \leq p \leq 0.5, (R_1, R_2) \notin \mathcal{R}_{\text{Exact}}(p)\}$$

(17)

and

$$P_{out}^{\text{Approx}} = \Pr \{0 \leq p \leq 0.5, (R_1, R_2) \notin \mathcal{R}_{\text{Approx}}(p)\}$$

(18)
respectively. Here, $R_{\text{Exact}}(p)$ and $R_{\text{approx}}(p)$ define the achievable rate regions, determined by (12) and (14), respectively.

The outage probability calculation is multi-fold integrals by taking into consideration of the pdfs of all the wireless links [25], [40], [41], [52], [53]. The numerical method from [96] can be utilized for the computation of multi-fold integrals.

### D. Diversity Order and Optimal Relay Location

The diversity order can be reflected from the outage probability curves, which indicates the slope of the outage probability in the high SNR regime. The outage probability with LF can be written as

$$P_{\text{out}}^{\text{LF}} = (G_c)^{-G_d}, \quad (19)$$

where $G_c$ and $G_d$ denote the coding gain and diversity order [53], [97], [98], respectively.

- When the S-R link is represented by a BSC with time-invariant crossover probability $p$ [41] and the other wireless links are supposed to suffer from Rayleigh block fading, the diversity order of the system is one unless $p = 0$. The diversity order is the same in the two ways of outage probability calculation, expressed in (17) and (18).

- When all the wireless links encounter Rayleigh block fading, we can achieve second order diversity with LF relaying. Same conclusion can be drawn for DF relaying. However, LF relaying outperforms DF relaying in terms of outage probability [25], [53].

- When all the wireless links encounter Nakagami-$m$ block fading, the equivalent diversity order is in the form of

$$G_d = m_1 + \min(m_0, m_2), \quad (20)$$

where $m_0$, $m_1$, and $m_2$ are the shape factors for S-R, S-D, and R-D links, respectively. We add the word “equivalent” before “diversity order” is because that diversity order is supposed to be a positive integer. However, due to the fact that the shape factors of the Nakagami-$m$ fading can be any positive real numbers, the equivalent diversity order in (20) can also be any positive real numbers.

The optimal relay location refers to the position where the minimal outage probability can be achieved. The classical experiment setup is that source, relay, and destination are placed in a line with the relay between the source and the destination [25]. The optimal location is the midpoint if the received SNRs of the S-R and R-D links have the same probability distribution, taking into account of the path loss, geometric gain, and shadowing effect. The reason lies in that the contributions of the S-R and R-D links are balanced. However, in DF relaying, the optimal relay position is closer to the source if the received SNRs of the S-R and R-D links have the same probability distribution [25], [53]. The optimal relay position is closer to the destination when the S-R link is a BSC with time-invariant crossover probability and the other links are Rayleigh block fading channels. This is because the quality of the S-R link is fixed while the relay location is, while it improves the R-D link quality by moving the relay closer to the destination.

The impact of the spatial correlation between the S-D and R-D links was intensively studied in [52]. As expected, the outage probability increases as the increase of the spatial correlation. Full diversity order can be achieved as long as the spatial correlation coefficient is not equal to one. The spatial correlation makes the optimal relay position move towards the destination.

### E. Upper-Layer Integration

Integration of LF with medium access and network layers were investigated in [82], [83]. To be specific, the authors in [82], [83] proposed a centralized lossy forwarding-medium access control (CLF-MAC) scheme, which allows lossy links and employs JD in a coordinated WLAN environment. The message exchange process in the uplink transmission is described in Fig. 9. The whole process is composed of two parts, contention free period (CFP) and contention period (CP). During the CFP, the source DATA, relay DATA, and end-to-end ACK are scheduled in a time division multiple access (TDMA) manner. During the CP, a dedicated resource reservation (RR) frame is sent by the source to inform the destination about the next planned uplink transmission. Better performance in terms of throughput can be achieved compared to coordinated carrier sense multiple access with collision avoidance (CSMA/CA) by allowing lossy links and JD of multiple received message copies. Moreover, the proposed CLF-MAC works well in the presence of hidden stations since it can avoid collision of DATA frames.

Employing JD at the destination brings new problems in the upper layers. Cross-layer design by considering both LF and JD should be investigated to make it applicable to the existing and future communication networks.

### F. Single-source Multi-relay Scenario

The extension from three-node network to the scenario of single-source multi-relay network with direct S-D link is not straightforward regarding the theoretical performance limits. The supporting theory is source coding with multiple helpers. The challenge lies in that the exact achievable rate region for source coding with multiple helpers is unknown yet. An upper bound for the outage probability was derived by relaxing and
reducing the rate constraints from the viewpoint of the Slepian-Wolf theorem in [90]. Recently, the problem of source coding with multiple helpers was studied in terms of the inner and outer bounds of the rate-distortion region [56]. However, the inner bound and the outer bound are not tight in general.

It should be noted that for practical code implementation, the low-complexity coding-decoding chain in Section III-B can be directly applied to the single-source multi-relay systems.

In the single-source multi-relay scenario, relay selection is another hot topic with intensive investigation in the literatures. The relay selection problem can be classified into two major categories, i.e., single-relay selection (SRS) [99]–[103] and multiple-relay selection (MRS) [102], [104], [105]. Intuitively, MRS is expected to achieve performance gain over SRS in terms of outage probability and symbol error rate (SER). However, the gain is in the form of parallel shift of the outage and/or SER curves to the lower-average SNR side, but no sharper decay corresponding to larger diversity order can be expected. Different criteria are used for choosing the optimal relay(s), e.g., maximizing the worse received SNR [99], [102], [103], minimizing the SER [100], minimizing the average BEP [104], minimizing the outage probability [100], [101].

More detailed information on the relaying schemes and selection criterion which literatures assume can be found in Table IV. With the relay selection techniques, \((m + 1)\)th diversity order can be achieved for the single-source \(m\)-relay networks with direct link, for instance, based on the criterion of maximizing the worse received SNR [102]. Fairness, e.g., outage priority and accumulative power consumption based proportional fair scheduling, was taken into consideration in [106], [107], aiming at improved selection fairness among all the relays without any sacrifice of the outage performance. However, all the relay selection schemes in the literatures are based on either DF or AF relaying. Due to the advantages of LF over DF and AF relaying, better performance, for instance, in terms of outage probability, is expected for the relay selection over LF relay network.

V. MULTIPLE ACCESS RELAY CHANNEL

We focus on the MARC network in this section, where both orthogonal and non-orthogonal channel access are considered and compared. The block diagram of orthogonal and non-orthogonal MARCs is presented in Fig. 10. For the orthogonal MARC [71], [72], [108]–[110], one transmission cycle requires three time slots. Source A (denoted by \(U_A\)) broadcasts its binary uniformly distributed information sequence to the relay and destination in the first time slot. Source B (denoted by \(U_B\)) becomes active and broadcasts its binary uniformly distributed information sequence to the relay and destination in the second time slot. In the third time slot, the relay gets the estimates of the source sequences (denoted by \(\hat{U}_A\) and \(\hat{U}_B\), respectively), conducts bit-wise XOR on them (i.e., \(\hat{U}_{A\oplus B} = \hat{U}_A \oplus \hat{U}_B\)), encodes, and forwards the encoded sequence to the destination. For the non-orthogonal MARC [69], [73], one transmission cycle requires only two time slots, since the non-orthogonal MARC combines phase 1 and phase 2 of orthogonal MARC into one phase.

In publications [108]–[110], the authors focused on designing practical joint network-channel codes (JNCCs), which are tailored for such network setup, based on LDPC codes and turbo-like codes. The DF relaying protocol was adopted at the relay. Theoretical limits of such network were studied from the perspective of network information theory [71], [72], by using multi-terminal source coding with a helper. In [71], perfect S-R links and orthogonal transmission were assumed, while in [72] the authors relaxed the constraint and assumed that S-R links suffer Rayleigh block fading. The theoretical limit analyses become more complicated when allowing non-orthogonal transmission, the region for lossless recovery was obtained by using the sufficient condition in [69], [73]. In [69], [72], [73], LF relaying protocol is adopted at the relay. A brief summary on the selected publications on MARC can be found in Table V.

A. Operations at Relay

Similar to the one-way relay network, the relay shows no interests in the original information sequences from the sources in both orthogonal and non-orthogonal MARCs. The relay decodes the information sequences, conducts bit-wise XOR on the decoded sequences, and then re-encodes and transmits the encoded sequence to the destination. The sequence sent by the relay is regarded as a helper for assisting the destination to decode the information sequences from the sources during the JD procedures.

Nevertheless, the operations at the relay are slightly different for orthogonal and non-orthogonal transmission MARCs. In the orthogonal MARCs, the relay decodes the information sequences from the sources separately and then performs bit-wise XOR on the estimates. No inter-user interference exists in the orthogonal transmission. However, in the non-orthogonal MARCs, the relay first implements multiuser detection (MUD) and then conducts bit-wise XOR on the estimates of the information sequences from the sources. The whole process is named as MUD-XOR in [73], [111].

With the application of LF relaying\(^4\), the relay re-encodes XORed version of the estimates, and sends it to the destination regardless of whether the estimates are correct or not. Error detection is not needed to operate at the relay with the application of LF relaying.

B. Achievable Rate Region

For the orthogonal transmission, the Shannon’s lossy source-channel separation theorem holds for each transmission link. The achievable rate region is obtained by following the theorem of correlated source coding with a helper [71], [72]. However, for the non-orthogonal transmission, the separation theorem does not hold. The computation of distortion would be a challenging task for the multiple access channel (MAC) consisting of the S-R links. Also, the exact region for lossless recovery over non-orthogonal MARC is unknown in general. In [69], [73], the region for lossless recovery was based on

\(^4\)However, with the application of DF, the relay will be silent if errors are detected either for source A or source B.
TABLE IV: SELECTED PUBLICATIONS ON RELAY SELECTION

<table>
<thead>
<tr>
<th>Article</th>
<th>Selection Criterion</th>
<th>Relaying Protocol</th>
<th>Channel Access Method</th>
<th>Performance Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>[99], [102], [103]</td>
<td>Best worse channel</td>
<td>AF and DF</td>
<td>Orthogonal</td>
<td>FER, SER, outage probability, and power efficiency</td>
</tr>
<tr>
<td>[100]</td>
<td>SER minimization</td>
<td>AF</td>
<td>Orthogonal</td>
<td>Throughput and outage probability</td>
</tr>
<tr>
<td>[104]</td>
<td>Average BEP minimization</td>
<td>AF</td>
<td>Orthogonal</td>
<td>Average BEP</td>
</tr>
<tr>
<td>[100], [101]</td>
<td>Outage probability minimization</td>
<td>AF</td>
<td>Orthogonal</td>
<td>Throughput and outage probability</td>
</tr>
</tbody>
</table>

---

![Diagram](a)  
![Diagram](b)

Fig. 10: The block diagram of MARC (a) orthogonal transmission, and (b) non-orthogonal transmission.

TABLE V: SELECTED PUBLICATIONS ON MARC WITH DF VS. LF, ORTHOGONAL VS. NON-ORTHOGONAL

<table>
<thead>
<tr>
<th>Article</th>
<th>Channel Type</th>
<th>Relaying Protocol</th>
<th>Channel Access Method</th>
<th>Performance Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>[109]</td>
<td>Rayleigh block fading</td>
<td>DF</td>
<td>Orthogonal and non-orthogonal</td>
<td>FER</td>
</tr>
<tr>
<td>[71]</td>
<td>Error-free S-R links and Rayleigh block fading for the remaining channels</td>
<td>DF</td>
<td>Orthogonal</td>
<td>FER, BER, and outage probability</td>
</tr>
<tr>
<td>[72]</td>
<td>Rayleigh block fading</td>
<td>LF</td>
<td>Orthogonal</td>
<td>FER and outage probability</td>
</tr>
<tr>
<td>[69], [73]</td>
<td>Rayleigh block fading</td>
<td>LF</td>
<td>Non-orthogonal</td>
<td>FER and outage probability</td>
</tr>
</tbody>
</table>

---

The sufficient condition, which combines the MAC capacity region and the Slepian-Wolf rate region [87, Ch. 14.1].

1) Orthogonal Transmission: For the orthogonal transmission, the achievable rate region is obtained by following correlated source coding with a helper [71], [72], i.e.,

\[ R_A \geq H(U_A|U_B) \]
\[ R_B \geq H(U_B|U_A) \]
\[ R_{A,B} = H(U_A, U_B) \]
\[ R_c \geq H(\hat{U}_{A\oplus B} | \hat{U}_{A\oplus B}) \]

where \( R_A, R_B, \) and \( R_c \) are the rates for \( U_A, U_B, \) and \( \hat{U}_{A\oplus B} \), respectively; \( \hat{U}_{A\oplus B} \) is the decoded version of \( U_{A\oplus B} \) at the destination. Compared to the achievable rate region determined by correlated source coding, the achievable rate region determined by the first three rows in (21) is enlarged because of the fact that conditioning reduces entropy [71].

The relationship between the transmission rates, distortion, and received SNR can be built by the Shannon’s lossy source-channel separation theorem,

\[ R_{c,A}(p_{A}) \leq C(\gamma_{A,R}) \]
\[ R_{c,B}(p_{B}) \leq C(\gamma_{B,R}) \]
\[ R_{c,R}(p_{R}) \leq C(\gamma_{R,D}) \]

where \( R(p_i) = 1 - H_2(p_i) \) for \( i \in \{A, B, R\} \) with \( p_A, p_B, \) and \( p_R \) being the distortion occurred in the source A-to-relay, source B-to-relay, and R-D links, respectively; \( R_{c,A}, R_{c,B}, \) and \( R_{c,R} \) are the normalized spectrum efficiency at sources A, B, and R, respectively; \( \gamma_{A,R}, \gamma_{B,R}, \gamma_{R,D} \) denote the instantaneous received SNR for the source A-to-relay, source B-to-relay, and R-D links, respectively.

After several manipulations, the achievable rate region described in (21) can be further expressed in

\[ R_{A,A} \geq H_2(p_A * p_B * p_R) \]
\[ R_{A,B} \geq H_2(p_A * p_B * p_R) \]
\[ R_{A,A} + R_{A,B} \geq 1 + H_2(p_A * p_B * p_R) \]

Because the destination does not allow any distortion, the S-D links are determined by the Shannon’s lossless source-channel separation theorem. The relationship among the compression rate, transmission rate (i.e., normalized spectrum efficiency), and instantaneous received SNR is given by,

\[ R_{c,A,R} \leq C(\gamma_{A,D}) \]
\[ R_{c,B,R} \leq C(\gamma_{B,D}) \]

with \( \gamma_{A,D} \) and \( \gamma_{B,D} \) being the instantaneous received SNR for the source A-to-destination and source B-to-destination links, respectively.

Taking (22)–(29) into consideration, successful transmission of both sources with lossless recovery at the destination can

---

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2018.2866711, IEEE Communications Surveys & Tutorials
be guaranteed if the following conditions hold [72]
\[
C(\gamma_{A,D})/R_{c,A} \geq H_2(p_a*p_B*p_R),
\]
\[
C(\gamma_{B,D})/R_{c,B} \geq H_2(p_a*p_B*p_R),
\]
\[
C(\gamma_{A,D}+\gamma_{B,D})/R_{c,B} \geq 1+H_2(p_a*p_B*p_R).
\]
(30) 
(31) 
(32)

2) Non-orthogonal Transmission: The difference between non-orthogonal transmission and orthogonal transmission is that it is extremely challenging to find the sufficient and necessary conditions for the lossless transmission over non-orthogonal MARCs. In [69], [73], we obtained the region for lossless recovery based on the sufficient condition, which combines the MAC capacity region and the Slepian-Wolf rate region [87, Ch. 14.1]. For the sake of simplicity, all the transmission rates are set to be equal, i.e., \( R_{c,A} = R_{c,B} = R_c \). Then, the region for lossless recovery at the destination for both of the sources are determined by [73]
\[
C(\gamma_{A,D})/R_c \geq H_2(p_e*p_R),
\]
\[
C(\gamma_{B,D})/R_c \geq H_2(p_e*p_R),
\]
\[
C(\gamma_{A,D}+\gamma_{B,D})/R_c \geq 1+H_2(p_e*p_R).
\]
(33) 
(34) 
(35)

The details of calculating \( p_e \) are described in [73] with \( p_e \) being the distortion occurred in the S-R links.

Due to the existence of non-orthogonal MAC composed of the S-R links, two different methods were taken into consideration when calculating the distortion. In [69], a virtual point-to-point channel was proposed to represent the non-orthogonal MAC. Then, the distortion can be derived using the Shannon’s lossy source-channel separation theorem. Alternative way is to rely on the MAC capacity region, the distortion can be computed by exploiting successive interference cancellation (SIC) in certain subregions [73].

C. Outage Probability

The outage probabilities for the orthogonal and non-orthogonal MARCs with LF relaying are derived in the same way as those for the first and second exemplifying scenarios. For the orthogonal MARCs, an outage event occurs if the rate pair \((R_{c,A},R_{c,B})\) falls outside the achievable rate region, defined by (21). As specified in (28) and (29), relationship among the instantaneous SNRs, compression rate, and transmission rate is built for the S-D links by following the Shannon’s lossless source-channel separation theorem. The achievable rate region can be further expressed by (30)–(32). The outage probability is obtained by calculating the probability of the region defined by (30)–(32) (i.e., \( P_{\text{out}}^{\text{LF}} = 1-Pr[\text{region defined by (30)–(32)}] \)) while taking into consideration the pdfs of the received SNRs for all the links. More details on the outage probability of orthogonal MARCs can be found in [72]. Similarly, the outage probability for the non-orthogonal MARCs with LF is determined by the region defined in (33)–(35) (i.e., \( P_{\text{out}}^{\text{LF}} = 1-Pr[\text{region defined by (33)–(35)}] \)). With \( C(\gamma_{A,D}+\gamma_{B,D}) < C(\gamma_{A,D}) + C(\gamma_{B,D}) \), in general, the outage probability of non-orthogonal MARC is larger than that of orthogonal MARC when \( R_{c,A} = R_{c,B} = R_c \). Nevertheless, the throughput of non-orthogonal MARC was found to be much higher than that of orthogonal MARC due to the reduced time slot consumption for one transmission cycle [73].

D. JD at the Destination

With respect to the practical code implementation for MARCs, we only focus on JD at the destination with ACC aided turbo codes applied at the transmitter side. The JD process also consists of two kinds of iterations, i.e., GI and LI. During the GI, the information is exchanged among different decoders, i.e., decoders for the two sources and decoder for the relay, by considering the correlation information between the sources (termed as source-source correlation) and the extrinsic information from the other decoders.

When the S-R links are error-free, the information exchange during GI can be written as [71]
\[
L_a(u^k_R) = \ln \left( \frac{1-p_u}{p_u} \right) + \ln \left( 1 + \exp[L_e(u^k_A) + L_e(u^k_B)] / \exp[L_e(u^k_A)] + \exp[L_e(u^k_B)] \right),
\]
\[
L_a(u^k_A) = \ln \left( \frac{p_u + (1-p_u) \exp[L_e(u^k_A) + L_e(u^k_B)]}{(1-p_u) \exp[L_e(u^k_A)] + p_u \exp[L_e(u^k_B)]} \right),
\]
\[
L_a(u^k_B) = \ln \left( \frac{p_u + (1-p_u) \exp[L_e(u^k_A) + L_e(u^k_B)]}{(1-p_u) \exp[L_e(u^k_A)] + p_u \exp[L_e(u^k_B)]} \right),
\]
(36) 
(37) 
(38)

where \( L_a(u^k_A), L_a(u^k_B) \), and \( L_a(u^k_R) \) denote the a priori LLRs for the ACC decoders of sources A and B, and relay with \( u^k_A, u^k_B, \) and \( u^k_R \) denoting the \( k \)th element of sequences \( U_A, U_B \) and \( U_R \), respectively. \( L_e(u^k_A), L_e(u^k_B), \) and \( L_e(u^k_R) \) denote the extrinsic LLRs for the ACC decoders of sources A and B, respectively, and \( p_u \) stands for the bit-flipping probability between the two sources, equivalent to source-source correlation.

When there exist intra-link errors, the information exchange among the three decoders is modified into \((39)–(41)\) according to [73], where \( p_{\text{out}} \) is the estimate of \( p_e \) at the destination during each GI, and \( p_{\text{e}} \) is the estimate of \( p_e \) at the destination.

When non-orthogonal transmission is assumed, one additional operation is required during the GI. We need to apply the modified demapping rule [73], [112], [113] to calculate the a priori LLRs from the channel for the two ACC decoders. The demapping rule is described by (42) and (43), where \( L_a(x^k_{A}) \) and \( L_a(x^k_{B}) \) are the a priori LLRs for the ACC decoders of sources A and B with \( x^k_{A} \) and \( x^k_{B} \) being the coded and modulated signal from sources A and B, respectively; \( L_e(x^k_{A}), L_e(x^k_{B}) \), and \( L_e(x^k_{R}) \) are the extrinsic LLRs for the ACC decoders of sources A and B, respectively, \( h_{A,D} \) and \( h_{B,D} \) are the channel coefficients for the source A-to-destination and source B-to-destination links, \( P_i \) stands for the transmit power at the sources, \( G_{i,j} \) is the geometric gain between node i and node j, \( r \in \{A,B\} \) and \( j \in \{D\} \), \( y_{0,1} \) is the received signal vector at the destination in the first time slot [73].

\(^5\)Note that the information exchange is expressed in symbol wise here while it is expressed in frame wise in Section III-B. In essence, they are identical.

\(^6\)Here, \( U_A, U_B \) and \( U_R \) are the realizations of \( U_A, U_B, U_{A+B} \), respectively.

\(^7\)For the orthogonal MARC, \( p_{\text{e}} \) is equal to \( p_a + p_{\text{out}} \). For the non-orthogonal MARC, the definition of \( p_{\text{e}} \) is the same as that defined in Section V-B2.
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2018.2866711, IEEE Communications Surveys & Tutorials

With the recognition of its usefulness, the two-way relay network has been intensively investigated with the aim of its applications to wireless communications [74], [114]–[116]. For example, two sources intend to ensure their information exchange via a common intermediate relay node in addition to the direct links between them, as illustrated in Fig. 11. With half-duplexing orthogonal transmission, three time slots are required for one transmission cycle. The literatures can be classified into different major groups based on their topic categories, e.g., rate region and performance bound analyses, practical limit-approaching coding/decoding design, relaying strategies in half- and full-duplex mode, and fairness issue.

The inner and outer bounds of the achievable rate region for the two-way relay network without direct link were first investigated in [117] and further studied in [114], assuming no cooperation between the two sources. The authors in [114] showed that cut-set upper bound can be achieved by using block Markov superposition coding together with Wyner-Ziv coding. As pointed in [116], three major protocols are applicable in such a network with DF, i.e., multiple access broadcast (MABC), time division broadcast (TDBC), and hybrid broadcast (HBC). A tight rate bound is derived for the MABC protocol. However, inner and outer bounds are not consistent with each other in the latter two protocols.

Physical layer network coding [115] is a promising technique to handle the interference at the relay node, simplify the decoding complexity, and enhance the network throughput by reducing the required transmission time slots. In practice, such operation can be performed independently or jointly with channel coding [17], [118]–[121]. For the practical limit-approaching schemes, it was shown in [122] that linear codes with subspace sharing are potential candidates for the binary case while lattice codes are asymptotically optimal for the Gaussian case in the high SNR regime.

Full duplex is another effective way to improve the spectrum efficiency by allowing the relay even all the nodes to transmit and receive on the same carrier frequency at the same time [123]–[126]. Therefore, the full-duplex nodes suffer from loop interference, which can be mitigated by advanced interference cancellation and transmit/receive antenna isolation [127]. Full-duplex two-way relay network (FD-TWRN) without direct link was investigated in [123]–[125] in terms of average rate and/or outage probability with the aid of physical layer network coding at the relay. Both AF and DF relaying were considered and compared under the assumption of perfect loop interference cancellation. Unlike [123]–[125], imperfect loop interference and SIC were taken into consideration based on DF relaying in [126]. FD-TWRN significantly outperforms half-duplex two-way relay network (HD-TWRN) [128], [129] in terms of achievable rates when the loop interference is below a certain threshold.

Applying LF relaying to two-way relay network with direct link was initially studied in [74] under the assumption that the relay works in half-duplex mode while the sources work in full-duplex mode. The first time-slot transmission is a two-user MAC, while the second time-slot transmission is the broadcast
of XoRed copy from the relay. The two-way relay channel is degraded into two correlated one-way relay network. Therefore, the achievable rate region can be formulated by source coding with a helper. The outage probability is derived to show its superiority over its DF counterpart in [74].

Selected publications on the two-way relay network classified based on half- and/or full-duplex relaying strategies are provided in Table VI.

Data rate fairness constraint assumes that the rates of both sources are identical in the information exchange. Optimal power allocation for the sources and relay, which leads to sum rate maximization, was studied in [130], [131] with AF and DF relaying. The extension to outage performance fairness was considered in [132] over asymmetric two-way AF relay network, where only the statistical information other than instantaneous information of the channel is required for the power allocation scheme.

VII. GENERAL MULTI-SOURCE MULTI-RELAY NETWORK

We study a more general network and derive its theoretical performance limits with the application of LF relaying at the relays. The system model is provided in Fig. 12, which consists of two sources, two relays, and one destination without any direct links. In the first hop transmission, the two sources broadcast their information sequences to the relays simultaneously. In the second hop transmission, the relays decode the information sequences from both the sources, re-encode, and send them to the destination. The end-to-end outage probability of such network was studied in [76], where all the links are assumed to suffer from i.i.d. Rayleigh block fading. Extension to i.i.d. $\kappa - \mu$ block fading was investigated in [75].

A. First Hop Transmission

The first hop is characterized by MAC capacity region [76]. There exist three possible outcomes for source 1’s decoded sequence at the two relays, provided in Table VII, where $\hat{u}_{1,R_1}$ and $\hat{u}_{1,R_2}$ denote the estimates of $u_1$ at relays 1 and 2, respectively. In the first case, both relays fail in successfully recovering the source 1’s sequence at the relay. On the contrary, both relays successfully decode the sequence from source 1 at the relay in the second case. In the third case, only one of the relays succeeds in recovering source 1’s sequence while the other fails. Because of the topology symmetry of the network, there also exist three possible outcomes for source 2’s decoded data at the two relays. The outage probability calculation for the first phase can be found in [75], [76] based on the MAC capacity region.

B. Second Hop Transmission

The transmission schemes utilized for the second hop heavily rely on the decoding outcomes at the two relays. We only concentrate on the forwarding of source 1’s decoded information sequence in the second hop transmission. For Case 1, the decoded sequences are erroneous at the relays. The transmission in the second hop becomes a chief executive officer (CEO) problem. For Case 2, maximum ratio transmission (MRT) can be applied at the relays to achieve the maximum diversity order. However, global CSI is required at the relays, which cannot be perfectly available in practical systems. The advantage lies in that the second hop transmission only needs one time slot. Alternatively, orthogonal transmissions can be applied in Case 2, where MRC is utilized at the destination.

In this regard, global CSI is required at the destination, which is relatively easier to achieve compared to CSI at the transmitter side. For Case 3, orthogonal transmission of the decoded sequences should be taken into consideration [76]. The problem can be formulated by source coding with a helper, which is introduced in Section IV. The incorrectly decoded sequence at one relay can be considered as a helper for the correctly decoded sequence at the other relay with JD at the destination.

C. End-to-End Outage Probability

Based on the Bayes’ rule, the end-to-end outage probability for source 1 is derived by taking into consideration of both hops, written as [76]

$$P_{\text{out}}^{\text{E2E}} = P_{\text{out}}^{\text{1,1}} P_{\text{out}}^{\text{2,1}} + (1 - P_{\text{out}}^{\text{1,1}})(1 - P_{\text{out}}^{\text{2,1}})P_{\text{out}}(\text{Case } 2)$$
$$+ (P_{\text{out}}^{\text{1,1}} - 1)(1 - P_{\text{out}}^{\text{1,1}})(P_{\text{out}}^{\text{2,1}})P_{\text{out}}(\text{Case } 3),$$

where $P_{\text{out}}^{i,j}$ denotes the outage probability of transmission between source $j$ and relay $i$ and $P_{\text{out}}(\text{Case } i)$ denotes the outage probability of Case $i$ during the second-hop transmission, which can be found in [75], [76].

VIII. CONCLUSION

A. Concluding Remarks

In this tutorial paper, we have provided a comprehensive survey on the achievable rate regions and outage analyses
for different cooperative communication network scenarios with the LF strategy. We have exploited the theorem of source coding with a helper to obtain the achievable rate region of the three-node network while we have leveraged the Slepian-Wolf theorem as an approximation. The extension to correlated fading, including temporal correlation and spatial correlation, has also been thoroughly studied, with a special focus on the diversity order and coding gain. Source coding with multiple helpers remains a challenging problem in the performance limit analysis for multi-relay scenario. However, inner and outer bounds, accurate but not yet tight, have been derived in the literatures. Upper-layer integration with LF has also been studied to show better throughput performance. Orthogonal and non-orthogonal MARCs have been introduced and compared in terms of achievable rate region, outage probability, and JD. Two-way relay network with LF has also been studied in terms of the theoretical performance limits, different operation modes of the nodes, and practical coding/decoding approaches. Finally, more general two-source two-relay network with LF relaying has been studied, and its outage performance has been derived. It has been demonstrated that LF outperforms DF in general in terms of the outage probability, FER, and $\epsilon$–outage probability at the sacrifice of slightly increased power consumption and computational complexity at the destination.

### B. Future Research Directions

The LF strategy has shown superior performance over its DF counterpart in terms of coding gain, outage probability, system coverage, etc., as described in the preceding sections. These merits enable the potential application of LF strategy to the existing and future cellular systems, e.g., LTE-A, 5G and beyond, as well as WSNs, and IoT networks. It is expected that with LF strategy, the overall performance of wireless networks such as throughput can be significantly improved to satisfy the increasing quality of service (QoS) requirements, as well as to boost network innovations. However, to enjoy the benefits of LF strategy, great effort has to be made on practical deployment and cross-layer design in order to make it compatible with the current network protocols. In this section, we highlight some of the future research directions which have not yet been fully investigated yet.

(a) **Information Theoretical Analysis**

Distributed lossless source coding theorems, e.g., Slepian-Wolf theorem and distributed lossless source coding with a helper theorem, play a crucial role in deriving the performance limits of some specified network topologies such as one-way relay network. However, for the networks having more complicated topologies, only generic bounds such as cut-set bound has been used, and no analytical approaches specific to the topologies can be used to perform theoretical analysis.

One example is single-source multi-relay network which corresponds to the source coding with multiple helper problem in network information theory. Since the achievable rate region of source coding with multiple helpers is still an open problem, the exact performance limits of single-source multi-relay network have not yet been obtained. The outage performance either relies on the inner or outer bounds of the achievable rate region. This motivates us to find tighter inner and outer bounds, which are expected to almost overlap with each other.

If all the relays cannot losslessly recover the original data from the source and there is no direct link between the source and the destination (e.g., Case 3 in the exemplifying scenario (e)), it becomes a binary CEO problem which is another unsolved problem in network information theory, except for some specific cases [133]. Binary CEO problem has attracted tremendous research interests. To derive the performance limits, more investigations on this problem such as tighter inner and outer bounds of the achievable rate-distortion region, should be carried out.

(b) **Physical Layer Security with LF**

Recently, physical layer security where secure data transmission from the source to the legitimate destination is based on the information-theoretical approaches has received significant research interests. The key idea of physical layer security is to degrade the channel between the source and the eavesdropper (referred to as wiretap channel) such that the channel capacity is less than that of the channel between the source and the destination (referred to as main channel). By selecting an appropriate coding rate at the source, i.e., the coding rate is larger than the channel capacity of the wiretap channel but smaller than the channel capacity of the main channel, secure transmission can be guaranteed according to Shannon’s channel coding theorem.

In the context of relay networks, physical layer security can be achieved by broadcasting artificial noise (AN) at the relay or destination to minimize the channel capacity of the wiretap channel, while AN can be efficiently eliminated at the destination. The drawback of AN is that

---

**TABLE VI: SELECTED PUBLICATIONS ON HALF- AND FULL-DUPLEX TWO-WAY RELAY NETWORK**

<table>
<thead>
<tr>
<th>Article</th>
<th>Duplex Mode</th>
<th>Relaying Protocol</th>
<th>Channel Access Method</th>
<th>Performance Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>[123]-[126]</td>
<td>Half duplex</td>
<td>AF and DF</td>
<td>Orthogonal and non-orthogonal</td>
<td>Sum rate, average rate, and outage probability</td>
</tr>
<tr>
<td>[128], [129]</td>
<td>Half duplex</td>
<td>AF</td>
<td>Orthogonal</td>
<td>Average mutual information and outage probability</td>
</tr>
<tr>
<td>[74]</td>
<td>Half duplex and full duplex</td>
<td>LF</td>
<td>Orthogonal and non-orthogonal</td>
<td>Outage probability</td>
</tr>
</tbody>
</table>

**TABLE VII: POSSIBLE DECODING OUTCOMES OF SOURCE 1’S INFORMATION SEQUENCE AT THE RELAYS**

<table>
<thead>
<tr>
<th>Case</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcomes</td>
<td>$\hat{u}<em>{1,R_1} \neq u_1, \hat{u}</em>{1,R_2} \neq u_1$</td>
<td>$\hat{u}<em>{1,R_1} = u_1, \hat{u}</em>{1,R_2} = u_1$</td>
<td>$(\hat{u}<em>{1,R_1} \neq u_1, \hat{u}</em>{1,R_2} = u_1)$ or $(\hat{u}<em>{1,R_1} = u_1, \hat{u}</em>{1,R_2} \neq u_1)$</td>
</tr>
</tbody>
</table>
stringent synchronization is required which is difficult in practice. The alternative way is transmit antenna selection (TAS) and/or relay selection in order to maximize the channel capacity of the main channel. However, with the concept of LF, it is possible for the eavesdropper to increase the channel capacity of the wiretap channel. Therefore, the design of efficient physical layer security framework as well as secrecy outage analysis in the presence of LF are challenging topics.

(c) Resource Allocation with LF
Evolved from mobile cloud computing, mobile edge computing (MEC) is considered as a promising technique in future 5G networks. In MEC, the computation and storage is pushed closer to the end user side by deploying physical servers at the edge devices, e.g., the base station. In this way, the transmission latency can be significantly reduced to meet the ultra-low latency requirement of 5G networks. The mobile end users (MEUs) are connected to the base station directly or via the relay nodes, and the radio resources are shared by the MEUs, the base station and the relay nodes.

With the LF strategy, it is expected that the channel assignment, power allocation between the MEUs and the relay station, and relay selection encounter new challenges. An efficient resource allocation scheme may facilitate the scheduling algorithm and interference management to avoid the transmission collisions which result in the throughput loss of the network. Since distributed algorithms are unable to achieve optimal resource allocation, centralized algorithms have to be more flexible by utilizing new networking techniques such as software defined networking (SDN). In this context, efficient channel assignment between different links, power allocation between the MEUs and the relay stations, and the best relay selection in the presence of LF strategy are worth investigating.

(d) Routing with LF
The routing protocol is of great importance to wireless networks, e.g., wireless sensor/mesh networks and mobile ad-hoc networks. The purpose of the routing protocol is to find the most appropriate path for each source node to establish a connection to its destination node. In wireless communications, to find the appropriate radio path, each node is assigned a routing metric that reflects the received signal strength, hop count or other performance related parameters, and the path with maximum or minimum routing metric is selected.

None of the existing routing algorithms, such as optimized link state routing (OLSR), dynamic source routing (DSR) and ad hoc on-demand distance vector (AODV), takes into account the characteristics of LF. With LF strategy, the coverage area of relay is enlarged, or the power consumption of relay nodes can be reduced while maintaining the same coverage area. A fundamental problem that still needs to be solved in multi-hop wireless networks with LF is jointly performing relay selection, resource allocation, and routing, such that optimal performance gains in terms of network capacity, energy consumption, and QoS can be achieved.
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2018.2866711, IEEE Communications Surveys & Tutorials


Jiguang He (S’16) received the B.Eng. degree in communications engineering from Harbin Institute of Technology, Harbin, China, in 2010, and M.Sc. degree also in communications engineering from Xiamen University, Xiamen, China, in 2013. From September 2013 to March 2015, he was with Key Laboratory of Millimeter Waves at City University of Hong Kong, conducting research on channel tracking over millimeter wave MIMO systems. Since June 2015, he has been with Centre for Wireless Communications (CWC), Department of Communications Engineering, University of Oulu, Oulu, Finland, studying toward his Ph.D degree. His research interests span cooperative communications, network information theory, joint source and channel coding, distributed source coding, and distributed compressive sensing.

Valteri Tervo (S’10-M’15) received the Dr.Sc. (Tech.) degree in telecommunications from the University of Oulu, Oulu, Finland, in 2015. He pursued the doctoral degree under the dual degree program between the University of Oulu and Japan Advanced Institute of Science and Technology (JAIST) which resulted in a Ph.D. degree from JAIST in 2015. He was the recipient of the Outstanding Performance Award from JAIST.

During 2015-2017, Dr. Tervo was working in Centre for Wireless Communications (CWC), University of Oulu, as a post-doctoral researcher and project manager. He was a Work Package Leader in FP7 project Links-on-the-fly Technology for Robust, Efficient, and Smart Communication in Unpredictable Environments (RESCUE). His research interests include power allocation and transceiver design for broadband wireless communications with special emphasis on iterative reception, physical and lower MAC layer algorithms, including error correction code design and optimization, in lossy decode-and-forward relay networks.

Currently, Dr. Tervo works as a specification project manager in Nokia, Finland.

Xiaobo Zhou (S’11-M’13) received the B.Sc. in electronic information science and technology from the University of Science and Technology of China (USTC), Hefei, China, the M.E. in computer application technology from Graduate University of Chinese Academy of Science (GUCAS), Beijing, China, and the Ph.D. degree from School of Information Science, Japan Advanced Institute of Science and Technology (JAIST), Ishikawa, Japan, in 2007, 2010, and 2013, respectively. From April 2014 to March 2015, he was a Researcher with the Department of Communications Engineering, University of Oulu, Oulu, Finland. Currently, he is an Associate Professor with the School of Computer Science and Technology, Tianjin University, Tianjin, China. His research interests include cooperative wireless communications, network information theory, cloud computing, data center networks, software-defined networking and mobile edge computing.

Xin He received the M.Sc. in Information Science from School of Information Science, Japan Advanced Institute of Science and Technology (JAIST), Ishikawa, Japan, in 2013, and the Ph.D. degree from JAIST and the University of Oulu, Oulu, Finland in 2016. He is currently a lecturer with the School of Computer and Information, Anhui Normal University. Also, he is a postdoc researcher with the School of Computer Science and Technology, University of Science and Technology of China (USTC), Hefei, China. His research interests include joint source-channel coding, cooperative wireless communications, network information theory, energy harvesting and backscatter communications.
Shen Qian graduated from the School of Telecommunication Engineering, Xidian University, Xi’an, China, in 2003. He received the M.Sc and Ph.D degrees in Information Science from Japan Advanced Institute of Science and Technology (JAIST) in 2014 and 2017, respectively. He received the second Ph.D degree in Communications Engineering from the University of Oulu, Finland in 2017. He was the recipient of the IEEE International Conference on Communications (ICC) Presentation Award (2016), IEICE Outstanding Student Award (2014), and NEC C&C Grants for Non-Japanese Researchers. He has been member of the Technical Program Committee (TPC) of the 2015 IEEE International Conference on Communications (ICC) CROWD-Net workshop, and the 2015 IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD).

Meng Cheng received the B.Eng. degree in telecommunication engineering from Anhui University of Technology (AHUT), Anhui, China, in 2009, the M.Sc. degree in wireless communications (with distinction) from the University of Southampton, Southampton, U.K., in 2010, and the Ph.D. degree in information science from Japan Advanced Institute of Science and Technology (JAIST), Ishikawa, Japan, in 2014. After that, he worked as a 5G research engineer in Shanghai Research Center of Huawei Technology Co., Ltd. In 2017, he returned to JAIST as a researcher. His research interests are network information theory, non-orthogonal multiple access, iterative receiver design and wireless geolocation.

Markku Juntti (S’93-M’98-SM’04) received his M.Sc. (EE) and Dr.Sc. (EE) degrees from University of Oulu, Oulu, Finland in 1993 and 1997, respectively.

Dr. Juntti was with University of Oulu in 1992–98. In academic year 1994–95, he was a Visiting Scholar at Rice University, Houston, Texas. In 1999–2000, he was a Senior Specialist with Nokia Networks. Dr. Juntti has been a professor of communications engineering since 2000 at University of Oulu, Centre for Wireless Communications (CWC), where he leads the Communications Signal Processing (CSP) Research Group. He also serves as Head of CWC – Radio Technologies (RT) Research Unit. His research interests include signal processing for wireless networks as well as communication and information theory. He is an author or co-author in almost 500 papers published in international journals and conference records as well as in books WCDMA for UMTS, Signal Processing Handbook and 5G Wireless Technologies. Dr. Juntti is also an Adjunct Professor at Department of Electrical and Computer Engineering, Rice University, Houston, Texas, USA.

Dr. Juntti is an Editor of IEEE TRANSACTIONS ON COMMUNICATIONS and was an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY in 2002–2008. He was Secretary of IEEE Communication Society Finland Chapter in 1996–97 and the Chairman for years 2000–01. He has been Secretary of the Technical Program Committee (TPC) of the 2001 IEEE International Conference on Communications (ICC 2001), and the Co-Chair of the Technical Program Committee of 2004 Nordic Radio Symposium and 2006 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2006), and the General Chair of 2011 IEEE Communication Theory Workshop (CTW 2011). He has served as Co-Chair of the Signal Processing for Communications Symposium of Globecom 2014 Signal Processing for Communications Symposium, IEEE GlobalSIP 2016 Symposium on Transceivers and Signal Processing for 5G Wireless and mm-Wave Systems, and ACM NanoCom 2018.

Tad Matsumoto (S’84-M’98-F’10) received his B.S., M.S., degrees, both under the mentorship by Prof. Shin-Ichi Takahashi, and Ph.D. degree under the supervision by Prof. Masao Nakagawa, from Keio University, Yokohama, Japan, in 1978, 1980, and 1991, respectively, all in electrical engineering.

He joined Nippon Telegraph and Telephone Corporation (NTT) in April 1980. Since he engaged in NTT, he was involved in a lot of research and development projects, all for mobile wireless communications systems. In July 1992, he transferred to NTT DoCoMo where he researched Code-Division Multiple-Access techniques for Mobile Communication Systems. In April 1994, he transferred to NTT America, where he served as a Senior Technical Advisor of a joint project between NTT and NEXTEL Communications. In March 1996, he returned to NTT DoCoMo where he served as a Head of the Radio Signal Processing Laboratory until August of 2001. He worked on adaptive signal processing, multiple-input multiple-output turbo signal detection, interference cancellation, and space-time coding techniques for broadband mobile communications. In March 2002, he moved to University of Oulu, Finland, where he served as a Professor at Centre for Wireless Communications. In 2006, he served as a Visiting Professor at Ilmenau University of Technology, Ilmenau, Germany, funded by the German MERCATOR Visiting Professorship Program. Since April 2007, he has been serving as a Professor at Japan Advanced Institute of Science and Technology (JAIST), Japan, while also keeping a cross-appointment position at University of Oulu.

He has led a lot of projects funded by Academy-of-Finland, European FP7, and Japan Society for the Promotion of Science as well as by Japanese private companies.

Prof. Matsumoto has been appointed as a Finland Distinguished Professor for a period from January 2008 to December 2012, funded by the Finnish National Technology Agency (Tekes) and Finnish Academy, under which he preserves the rights to participate in and apply to European and Finnish national projects. Prof. Matsumoto is a recipient of IEEE VTS Outstanding Service Award (2001), Nokia Foundation Visiting Fellow Scholarship Award (2002), IEEE Japan Council Award for Distinguished Service to the Society (2006), IEEE Vehicular Technology Society James R. Evans Avant Garde Award (2006), and Thuringen State Research Award for Advanced Applied Science (2006), 2007 Best Paper Award of Institute of Electrical, Communication, and Information Engineers of Japan (2008), Telecom System Technology Award by the Telecommunications Advancement Foundation (2009), IEEE Communication Letters Exemplary Reviewer (2011), Nikkei Wireless Japan Award (2011), IEEE VTS Recognition for Outstanding Distinguished Lecturer (2016), and IEEE Transactions on Communications Exemplary Reviewer (2018). He is a Fellow of IEEE and a Member of IEICE. He is serving as an IEEE Vehicular Technology Distinguished Speaker since July 2016.