Short-Block Length Polar-Coded Modulation for the Relay Channel

Heping Wan, *Member, IEEE*, and Aria Nosratinia, *Fellow, IEEE*

***Abstract*— Short-block length communication is becoming increasingly important for applications such as machine-to-machine communication, among others. This paper studies polar-coded modulation in the short-block length regime for decode-forward, amplify-forward and compress-forward relays. Our work seamlessly combines multi-level signaling, polar coding/decoding, and the requirements of relaying, into encoders and decoders that exhibit excellent performance. Compared with the state of the art in decode-forward, our work demonstrates 2.5 dB improvement (at block length 512). For shorter block lengths 256 and 128, this work is the first reported implementation for any coded modulation and any relaying protocol. In the category of polar-coded *full-duplex* relaying, this work presents the first implementation *at any block length* and for any relaying protocol. For decode-forward, we propose and analyze joint iterative belief propagation decoding with polar codes, and successive list decoding with polarization-adjusted convolutional (PAC) codes. For amplify-forward, we utilize PAC codes and successive list decoding. For the three relaying protocols, the respective dispersion bounds are presented for comparison, and the error exponent of the multi-level relaying coded modulations are analyzed to shed light on the results. Extensive simulations verify the performance of the proposed coding schemes.**

***Index Terms*— Low-latency reliable communication, short-block length, polar code, coded modulation, relay, decode-forward, amplify-forward, compress-forward.**

I. Introduction

RELIABLE low latency communication, realized by short-block length transmission, is in high demand. Ultra-low latency is part of the 5G wireless standard, e.g., to support remotely-controlled applications (e.g. remote surgery), Internet of Things, and machine-to-machine (M2M) communication [1], [2]. Some other applications of short-block length transmission are discussed in [1], [2], and [3].

In the short-block length regime, polar codes [4] are the leading candidate for error control. With the combination of successive list decoding and cyclic redundancy check (CRC) [5], polar coding was shown to approach Polyanskiy's dispersion bound [6]. Arıkan [7] proposed PAC codes consisting of a rate-1 convolutional code and polar code, which also approached Polyanskiy's dispersion bound, and has better performance than simple polar codes under extremely short-block lengths. To avoid the disadvantages of hard-decisions in list decoding of polar codes, in [8], [9], and [10], a belief propagation (list) decoding was proposed based on updating soft information iteratively over the graph implied by the polar transform matrix.

Operation in the high spectral efficiency regime requires coded modulation. Seidl et al. [11] jointly designed polar codes and multi-level modulation labeling rules, motivated by similarities between multistage decoding and polar decoding [12]. Dai et al. [13] investigated the performance of a certain class of multi-level polar-coded modulation related to the 5G standard [14]. Other works on multi-level polar-coded modulation include [15], [16], [17].

Relaying is a widely accepted technique for improving the performance in multi-node networks. Under short-block lengths, achieving low error rates is harder, making it natural to call on relaying for help. We briefly review the most relevant literature on coding and coded modulation for the relay channel. Rodriguez et al. [18] investigated the error performance and diversity behavior of amplify-forward relay using bit interleaved coded modulation (BICM) under the effect of residual self-interference. Wan et al. [19] proposed a multi-level compress-forward relaying with trellis coded quantization. Abotabl and Nosratinia [20] investigated a bit-additive superposition scheme for decode-forward relaying. Also noteworthy are [21], [22] for relaying in the context of physical layer network coding, and [23], [24], [25], [26] for low density parity check (LDPC) code design for relaying. Blasco-Serrano et al. [27] showed that nested polar codes are suitable for binary symmetric decode-forward and compress-forward relay channel with orthogonal receivers, and then Karas et al. [28] proposed a half-duplex decode-forward with block length 32 and 128. For compress-forward relaying under 16-QAM and block length 4096, Madhusudhanan and Nithyanandan [29] found that polar codes have superior error performance compared with Turbo codes.

Remark 1: The literature on relay coding is not always explicit about the meaning of block length, *which can refer either to the length of codeword transmitted from source to relay, or the length of the end-to-end block that consists of both*

Manuscript received 2 March 2022; revised 3 August 2022; accepted 28 October 2022. Date of publication 30 November 2022; date of current version 16 January 2023. This work was supported in part by the National Science Foundation under Grant 1711689 and Grant 2008684. An earlier version of this paper was presented in part at the IEEE Global Communications Conference (Globecom) 2021. The associate editor coordinating the review of this article and approving it for publication was I. Tal. *(Corresponding author: Aria Nosratinia.)*

The authors are with the Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: heping.wan@utdallas.edu; aria@utdallas.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2022.3225546.

Digital Object Identifier 10.1109/TCOMM.2022.3225546

TABLE I
LIST OF THE RESULTS OF THIS PAPER

	Source-relay block length	End-to-end block length	Modulation	Codes
Decode-forward	128	256	8-PSK	polar & PAC
	256	512	8-PSK	polar & PAC
	512	1024	16-QAM	polar
Amplify-forward	128	256	8-PSK	polar & PAC
	256	512	8-PSK	polar & PAC
Compress-forward	128	256	16-QAM	polar
	256	512	16-QAM	polar

the source-transmitted and relay-transmitted symbols. The former is more suitable for highlighting the smallest coding block lengths appearing in the context of a relay channel, while the latter is more suitable for comparing short-block length relays against a non-cooperative baseline. Throughout the remainder of this paper, block length *refers to the source-to-relay codeword. For clarity, the corresponding end-to-end block length for each of our results is also pointed out in Table I.*

The closest work to the results of this paper is [30], which studied half-duplex, orthogonal, decode-forward, multi-level polar-coded modulation at block length 512.[1]

This paper studies polar-coded modulation in the short-block length regime for multiple relaying protocols and block lengths (see Table I). Under decode-forward, we study multi-level polar codes, as well as multi-level PAC codes, with successive list decoding for PAC code and joint iterative belief propagation decoding for stand-alone[2] polar-coded modulation. We also present the results for a half-duplex version for comparison with earlier literature [30]. Our amplify-forward polar-coded modulation scheme has a source node utilizing PAC codes, and at the destination employs successive list decoding. For compress-forward relaying, we utilize a multi-level polar-coded modulation at the source and the relay. The relay compression is achieved by scalar quantization, and the destination employs joint iterative belief propagation decoding. Extensive simulations demonstrate the performance of our coding schemes. A list of simulation results of this paper is presented in Table. I. In summary, the contributions of this paper are as follows:

- We design, implement, and evaluate multi-level polar coded modulation for short-block length relaying, under decode-forward, compress-forward, and amplify-forward protocols.
- A joint decoder at the destination is implemented for multi-level polar-coded modulation, removing the necessity of binning-type encoding[3] at the relay. Joint decoders have been explored via random coding arguments, but this work contributes the first implementation and performance demonstration with structured coding/modulation.

[1]In [30], the corresponding end-to-end block length is specified, which is 1024.

[2]as opposed to concatenated polar-coded modulation, i.e., PAC code.

[3]E.g., Wyner-Ziv for compress-forward and universal polar codes, which have been used to show performance limits for decode-forward.

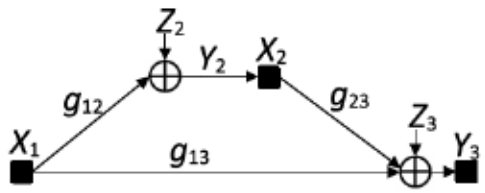


Fig. 1. Gaussian relay channel.

- We obtain 2.5 dB improvement at block-length 512 over the state of the art in half-duplex decode-forward.
- At block lengths 256 and 128, our work is the first report of successful implementation for *any relaying protocol under any coded modulation*.
- In the category of polar-coded *full-duplex* relaying, our work presents the first successful implementation *at any block length and for any relaying protocol.*
- We present an error exponent analysis for relay multi-level coded modulation, and calculate dispersion bounds.

II. System Model, Notation, and Preliminaries

In this paper, random values are represented with upper-case letters, e.g. Y_2, deterministic values with lower case letters, e.g. y_2, and corresponding vectors with bold font, e.g., $\boldsymbol{Y}_2$, $\boldsymbol{y}_2$. Sets are denoted with script letters, e.g., $\mathcal{A}$. Probability densities and mass functions are shown with $p(\cdot)$, and the distinction of discrete and continuous variables is clear from context. $P(\cdot)$ returns the probability of its argument (an event), and E denotes the error event.

In our full-duplex relay model, following [20], self-interference is modeled as additive white Gaussian noise (AWGN), and is absorbed together with the channel noise. The three-node AWGN full-duplex relay channel is shown in Fig. 1. The source and the relay, respectively, have transmit signals X_1, X_2, with average power constraints ρ_s, ρ_r. The AWGN, unit-variance receiver noise at the relay and destination are denoted Z_2, Z_3. The three channel coefficients g_{13}, g_{12} and g_{23} are known by all three nodes. The received signals are:

$$Y_2 = g_{12}X_1 + Z_2,$$
$$Y_3 = g_{13}X_1 + g_{23}X_2 + Z_3.$$

Full-duplex relaying is implemented by a block-Markov strategy in which a sequence of $b-1$ messages $\{d^{(j)}\}_1^{b-1}$is transmitted over b transmission blocks. By convention $d^{(0)} = d^{(b)} = 1$.

A. Polar Coding & Belief Propagation Decoding

Polar codes [4] are defined with a generator matrix $G_2^{\otimes n}$, where $\otimes$ is the Kronecker product and $G_2 \triangleq \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. The data to be encoded is denoted with $\boldsymbol{U} \triangleq [U_1, \cdots, U_N]$ where $N = 2^n$. The codeword is denoted with $\boldsymbol{x} = \boldsymbol{u}G_2^{\otimes n}$. The instances of the channel are denoted with $W : X \to Y$ and the polarization induces N virtual channels. We denote with $\mathcal{G}$

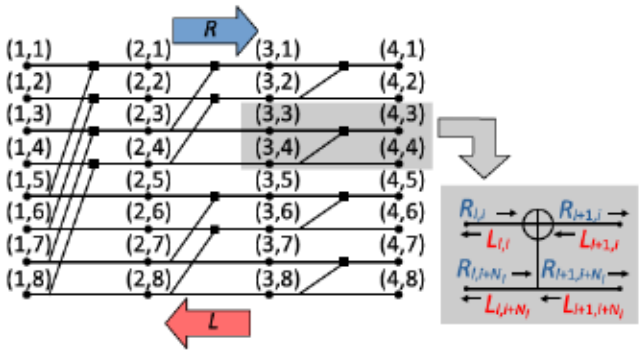


Fig. 2. Factor graph characterized by $\boldsymbol{h} = [1, 2, 3]$, and its processing unit.

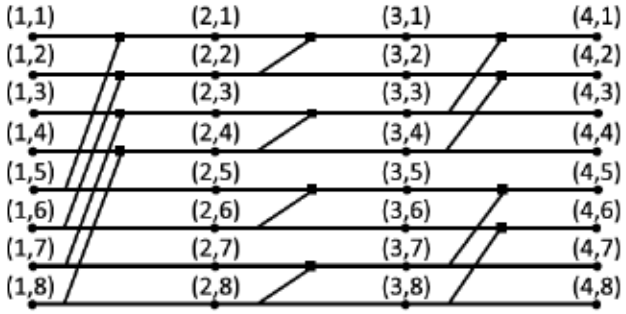


Fig. 3. Factor graph permutation characterized by $\boldsymbol{h} = [1, 3, 2]$.

the set of indices for the reliable virtual channels. Following Arıkan [4], we call $\mathcal{G}$ the *information set* and its complement $\mathcal{G}^c$ the *frozen set*. The message and frozen bits are carried by $\{U_\ell : \ell \in \mathcal{G}\}$ and $\{U_\ell : \ell \in \mathcal{G}^c\}$ respectively.

At the destination, two main families of decoding schemes including successive cancellation decoding and belief propagation decoding are applied to recover the estimated $\hat{\boldsymbol{u}}$ using the knowledge of $\{u_\ell : \ell \in \mathcal{G}^c\}$.

Belief propagation decoding is implemented over the original factor graph which is corresponding to $G_2^{\otimes n}$ and represented by the vector $\boldsymbol{h} = [1, 2, \cdots, n]$, for example Fig. 2 shows the original factor graph for $N = 8$. In the graph, each node labeled by (l, i) has a right-to-left message $L_{l,i}$ and left-to-right message $R_{l,i}$ updated iteratively via the basic processing unit (also shown in Fig. 2) containing the mapping:

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

$$L_{l,i+N_l} = f(L_{l+1,i}, R_{l,i}) + L_{l+1,i+N_l}, \quad (2)$$

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

$$R_{l+1,i+N_l} = f(L_{l+1,i}, R_{l,i}) + R_{l,i+N_l}, \quad (4)$$

where $1 \le l \le n+1$, $1 \le i \le N$, $N_l = 2^{n-h_l}$ and the function $f(x, y) = \ln \frac{1+e^{x+y}}{e^x+e^y}$.

$L_{n+1,i}$ is initialized to log-likelihood ratio (LLR) $\log\left(\frac{p(x_i=0|y_i)}{p(x_i=1|y_i)}\right)$, $R_{1,i}$ to ∞ if $i \in \mathcal{G}^c$, and the remaining messages to zero. The decoding is conducted by iteratively updating the right-to-left messages from stage n to 1 via Eqs. (1) and (2) and the left-to-right messages from stage 2 to $n+1$ via Eqs. (3) and (4). When the maximum number of iterations is met or an early stopping condition is satisfied, stop to produce the LLRs of the estimated $\hat{u}_i$ and $\hat{x}_i$ as $R_{1,i} + L_{1,i}$ and $R_{n+1,i} + L_{n+1,i}$ respectively.

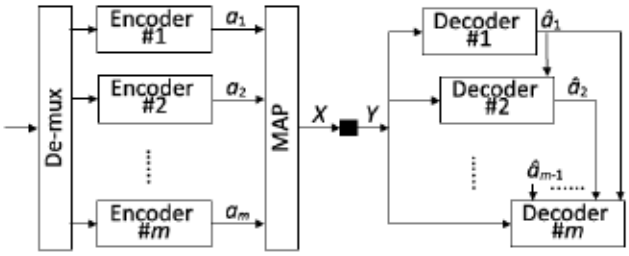


Fig. 4. Multi-level coding with multistage decoding in point-to-point channel.

A belief propagation list decoder [9] contains belief propagation decoders based on different permutations (e.g. Fig. 3) of the original factor graph. Component belief propagation decoders iteratively update the messages in parallel until the maximum number of iterations or a stopping condition is met. If a belief propagation decoder has the stopping condition satisfied and the estimated codeword closest to the channel observation in Euclidean distance among all the estimated codewords satisfying stopping condition, its output is declared to be the output of the list decoder.

B. Multi-Level Polar-Coded Modulation

Multi-level coding (see Fig. 4) demultiplexes the data into m bit-streams, each transmitted via one of the bits in the modulation label, and protected independently by a binary error-control code. At each time instance, the output of encoders denoted $\boldsymbol{A} \triangleq [A_1, \cdots, A_m]$ is mapped to the transmit signal X via a (bijective) mapping. Multistage decoding is motivated by the chain rule for mutual information:

$$I(X;Y) = I(\boldsymbol{A};Y) = \sum_{k=1}^{m} I(A_k;Y|A_1, \cdots, A_{k-1}), \quad (5)$$

where the bijective nature of mapping to modulation symbols has been used, and A_0 represents a constant. Successive decoding is then made possible in the sub-channels implied by Eq. (5), namely $W_k : A_k \to \{Y, A_1 \cdots A_{k-1}\}$. At each level k, the multistage decoder employs the channel observation y as well as the decoded values from preceding levels. To ensure that decoding at each level is reliable, the code rate at level k is chosen to be less than $I(A_k;Y|A_1, \cdots, A_{k-1})$. The constellation constrained capacity can be achieved by multi-level coding if and only if the optimal input distribution of the channel can be presented as the product of the marginal distributions of A_k [31].

Let $\boldsymbol{U}_k \triangleq [U_{k,1}, \cdots, U_{k,N}]$ and $\boldsymbol{A}_k \triangleq [A_{k,1}, \cdots, A_{k,N}]$ denote the data to be encoded and the codeword at level k. u_k is mapped to a_k via the polar transform $G_2^{\otimes n}$. Polarization of W_k induces mN virtual sub-channels denoted $W_{k,i} : U_{k,i} \to \{\boldsymbol{Y}, \boldsymbol{U}_1 \cdots \boldsymbol{U}_{k-1}, U_{k,1}, \cdots, U_{k,i-1}\}$. Density evolution with/without Gaussian approximation [32], [33] and Monte Carlo method [16] are used to determine the information sets and the frozen sets at all levels. Using the knowledge of frozen bits, the destination employs the multistage decoding with successive cancellation or belief propagation decoding at each level to recover the estimated $\hat{u}_k$.

III. Decode-Forward

A. Multi-Level Polar-Coded Modulation

The relay decodes the transmitted message by observing y_2, and helps the source by retransmitting the decoded message in the next block. The achievable rate is given by [34]:

$$R_{DF} \leq \max_{p(x_1,x_2)} \min\{I(X_1;Y_2|X_2),\, I(X_1,X_2;Y_3)\}. \quad (6)$$

In the following, we express this rate in the context of multi-level coding. The source transmit signal X_1 is derived using a bijective mapping from a binary vector $\boldsymbol{A} \triangleq [A_1,\ldots,A_m]$, and similarly the relay transmit signal X_2 has a bijective mapping with a binary vector $\boldsymbol{B} \triangleq [B_1,\ldots,B_m]$. Each element of $\boldsymbol{A},\boldsymbol{B}$ is the output of a binary error control code. For ease of exposition we consider the case where source and relay multi-level coding has the same number of levels; however, this is easily extended to modulations with different alphabet sizes [20].

By applying multi-level coding at the source and the relay, the achievable rate (6) can be expressed as

$$\begin{aligned} R_{DF} \leq \max \min \Bigg\{ & \sum_{k=1}^{m} I(A_k;Y_2|\boldsymbol{B},A_1,\cdots,A_{k-1}), \\ & \sum_{k=1}^{m} I(A_k;Y_3|\boldsymbol{B},A_1,\cdots,A_{k-1}) \\ & +I(B_k;Y_3|B_1,\cdots,B_{k-1}) \Bigg\}, \end{aligned} \quad (7)$$

where maximization is over $\prod_{k=1}^{m} p(a_k|b_k)p(b_k)$ which means the source signal at level k only depends on the relay signal at the same level. Multi-level coding is optimal if

$$\prod_{k=1}^{m} p^*(a_k|b_k)p^*(b_k) = p^*(\boldsymbol{a}|\boldsymbol{b})p^*(\boldsymbol{b}),$$

where $p^*(\cdot)$ represents the optimal distribution. The achievable rate at level k satisfies

$$\begin{aligned} R_{DF}(k) \leq \max \min \{ & I(A_k;Y_2|\boldsymbol{B},A_1,\cdots,A_{k-1}), \\ & I(A_k;Y_3|\boldsymbol{B},A_1,\cdots,A_{k-1}) \\ & +I(B_k;Y_3|B_1,\cdots,B_{k-1})\}. \end{aligned} \quad (8)$$

By introducing auxiliary random variables V_k, the achievable rate (8) is equivalent to [35]

$$\begin{aligned} R_{DF}(k) \leq \max \min \{ & I(V_k;Y_2,\boldsymbol{B}|V_1,\cdots,V_{k-1}), \\ & I(V_k;Y_3,\boldsymbol{B}|V_1,\cdots,V_{k-1}) \\ & +I(B_k;Y_3|B_1,\cdots,B_{k-1})\}, \end{aligned} \quad (9)$$

maximized over $\prod_{k=1}^{m} p(v_k)p(b_k)$ and functions $\{f_k : (v_k,b_k) \to a_k,\, k=1,\cdots,m\}$ so that V_k is uniform.[4]

[4]The conditional probabilities $p(a_k|b_k)$ that can be represented in this way are rational, however, with a large cardinality for V_k, any real-valued conditional probability can be approximated with arbitrary accuracy. This is sufficient, considering mutual information is a continuous function of the underlying distributions.

1) Codebooks: At each level k, the source utilizes a code that is characterized by a polarization transform H_k, the concatenation of the polar transform [4] and the universal polar transform [35]. At each level k, the source transmits into a binary broadcast channel. This broadcast channel has two point-to-point representations, namely $W'_k : V_k \to \{Y_2,\boldsymbol{B},V_1\cdots V_{k-1}\}$ for the source-relay link and $W''_k : V_k \to \{Y_3,\boldsymbol{B},V_1\cdots V_{k-1}\}$ for the source-destination link. Under H_k, the information set for W'_k is denoted $\mathcal{A}_k$, and the information set for W''_k is denoted $\mathcal{B}_k$. Assume that $I(W''_k) < R_{DF}(k)$, otherwise the message can be directly transmitted over the source-destination link. According to [35, *Theorem 1*], the following facts are true: First, $\lim_{N\to\infty} \frac{|\mathcal{A}_k|}{N} = R_{DF}(k)$. Second, $\lim_{N\to\infty} \frac{|\mathcal{B}_k|}{N} = I(W''_k)$. Third, $\mathcal{B}_k$ is a subset of $\mathcal{A}_k$.

Let $\{U_{k,\ell} : \ell \in \mathcal{A}_k\}$ carry the message bits at level k and $\{U_{k,\ell} : \ell \in \mathcal{A}_k\backslash\mathcal{B}_k\}$ represent its binning index. $U_{k,\ell} = u, \forall \ell \in \mathcal{A}_k^c$ are fixed and revealed to the relay and the destination. At the relay, the polar codes in Section II-B are adopted to send the binning index of the recovered message over the point-to-point sub-channels denoted $W'''_k : B_k \to \{Y_3,B_1\cdots B_{k-1}\}$. Because of $\lim_{N\to\infty} \frac{|\mathcal{A}_k\backslash\mathcal{B}_k|}{N} = R_{DF}(k) - I(W''_k) \leq I(W'''_k)$ observed from the achievable rate (9), the relay codewords are decodable at the destination with sufficientl large N. Let $\mathcal{I}_k$ denote the information sets corresponding to W'''_k and hence $|\mathcal{I}_k| = |\mathcal{A}_k\backslash\mathcal{B}_k|$. The bits whose indices belong to $\mathcal{I}_k^c$ are frozen and revealed to all three nodes.

2) Encoding and Decoding: In block j, the source demultiplexes the message $d^{(j)}$ into m bit-streams according to $R_{DF}(k)$, each carried by $\{u_{k,\ell}^{(j)} : \ell \in \mathcal{A}_k\}$. Computing $u_k^{(j)}H_k$ results in $v_k^{(j)}$. It then relocates $\{u_{k,\ell}^{(j-1)} : \ell \in \mathcal{A}_k\backslash\mathcal{B}_k\}$ to the bits whose indices belong to $\mathcal{I}_k$. Computing $u_k^{(j-1)}G_2^{\otimes n}$ results in $b_k^{(j-1)}$. Therefore, $a_{k,i}^{(j)} = f_k(v_{k,i}^{(j)}, b_{k,i}^{(j-1)})$. The source transmits $x_1(d^{(j)}|\omega^{(j-1)})$, where $\omega^{(j-1)} \triangleq \{u_{k,\ell}^{(j-1)} : \ell \in \mathcal{I}_k, k=1,\cdots,m\}$ denotes the binning index of $d^{(j-1)}$.

Knowing $\tilde{u}_{k,\ell}^{(j)} = u_{k,\ell}, \forall \ell \in \mathcal{A}_k^c$, the relay recovers an estimated $\{\bar{u}_{k,\ell}^{(j)} : \ell \in \mathcal{A}_k\}$ successively as

$$\bar{u}_{k,i} = \arg\max_{u\in\{0,1\}} p\big(u|y_2^{(j)}, x_2(\bar{\omega}^{(j-1)}),\, \bar{v}_1^{(j)}\cdots\bar{v}_{k-1}^{(j)}, \tilde{u}_{k,1},\cdots,\tilde{u}_{k,i-1}\big),$$

and transmits $x_2(\bar{\omega}^{(j)})$ in block $j+1$.

At the end of block $j+1$, with the knowledge of frozen bits whose indices belong to $\mathcal{I}_k^c$, the destination first recovers the estimated $\{\hat{u}_{k,\ell}^{(j)} : \ell \in \mathcal{I}_k\}$ (equivalently $\{\hat{u}_{k,\ell}^{(j)} : \ell \in \mathcal{A}_k\backslash\mathcal{B}_k\}$) based on the channel observation $y_3^{(j+1)}$ and the decoded codewords $\{\hat{b}_1^{(j)}\cdots\hat{b}_{k-1}^{(j)}\}$ from preceding levels. Knowing $\{\hat{u}_{k,\ell}^{(j)} : \ell \in \mathcal{A}_k\backslash\mathcal{B}_k\}$ and $\hat{u}_{k,\ell}^{(j)} = u_{k,\ell}, \forall \ell \in \mathcal{A}_k^c$, it then recovers $\{\hat{u}_{k,\ell}^{(j)} : \ell \in \mathcal{B}_k\}$ successively as

$$\hat{u}_{k,i} = \arg\max_{u\in\{0,1\}} p\big(u|y_3^{(j)}, x_2(\hat{\omega}^{(j-1)}),\, \hat{v}_1^{(j)}\cdots\hat{v}_{k-1}^{(j)}, \hat{u}_{k,1},\cdots,\hat{u}_{k,i-1}\big).$$

B. Design of Decode-Forward Polar-Coded Modulation

The universal polar codes for W_k' and W_k'' are impractical in the short-block length regime, due to the requirement of larger block length for channel polarization as well as higher complexity in encoding and decoding than the polar codes. Thus, our work employs the polar codes, i.e. $H_k = G_2^{\otimes n}$, at the source. The construction of polar codes at the source and the relay follows the ordered sequence in 5G standard [14] which is independent of the channels and has shown robust performance in multi-level polar-coded modulation [13].

At the relay, we transmit the entire estimated $\{\tilde{u}_{k,\ell}^{(j)} : \ell \in \mathcal{A}_k\}$ instead of its binning index $\{\tilde{u}_{k,\ell}^{(j)} : \ell \in \mathcal{A}_k \backslash \mathcal{B}_k\}$. In this case, $\mathcal{I}_k = \mathcal{A}_k$ because of the usage of the same ordered sequence in 5G. The relay transmits $x_2(\tilde{d}^{(j)})$ in block $j+1$. Since the rate of the relay codebook may be above the capacity of the relay-destination link, i.e. $\frac{|\mathcal{A}_k|}{N} > I(W_k''')$, the destination may not correctly decode the relay codewords. However, the relay codewords can be recovered jointly together with the source codewords [20].

Remark 2: Our polar encoder at the source generates codewords that depend only on the present message, while the full block-Markov encoding source has codewords that depend on both the present and past message. This simplification is equivalent to $a_k = v_k$ and calculating the achievable rates over the set of distributions $p(x_1)p(x_2)$ instead of $p(x_1, x_2)$. This was done in the interest of simplicity of code design as well as avoiding complications in decoding.

In the following, we propose two joint decoding algorithms used at the destination: joint iterative belief propagation decoding and successive cancellation decoding.

1) Joint Iterative Belief Propagation Decoding: At the end of block $j+1$, the joint decoder implemented through iterative belief propagation is based on $y_3^{(j+1)}$, the destination observation in block $j+1$, and $y_3^{(j)}$, the destination observation in block j from which the relay transmission $x_2(\hat{d}^{(j-1)})$ has been peeled off, i.e.,

$$\bar{y}_3^{(j)} = y_3^{(j)} - g_{23}x_2(\hat{d}^{(j-1)}) = g_{13}x_1(d^{(j)}) + z_3, \quad (10)$$

The destination searches for the candidate codewords $\big(x_1(\hat{d}^{(j)}), x_2(\hat{d}^{(j)})\big)$ that maximize the *a posteriori* probability $p\big(x_1(\hat{d}^{(j)})x_2(\hat{d}^{(j)})|\bar{y}_3^{(j)}y_3^{(j+1)}\big)$ which can be factorized as

$$p(x_1x_2|\bar{y}_3y_3) = \frac{1}{p(\bar{y}_3)p(y_3)}p(x_1x_2\bar{y}_3y_3), \quad (11)$$

where

$$\begin{aligned} &p(x_1x_2\bar{y}_3y_3) \\ &\overset{(a)}{=} p(\bar{y}_3|x_1)p(y_3|x_2)p(x_1)p(x_2), && (12)\\ &\overset{(b)}{\propto} p(\bar{y}_3|x_1)p(y_3|x_2)\mathbb{1}\{x_1 \in \mathcal{C}_\mathbf{S}\}\mathbb{1}\{x_2 \in \mathcal{C}_\mathbf{R}\}, && (13)\\ &\overset{(c)}{=} \prod_{i=1}^{N} p(\bar{y}_{3,i}|x_{1,i})p(y_{3,i}|x_{2,i})\mathbb{1}\{x_1 \in \mathcal{C}_\mathbf{S}\}\mathbb{1}\{x_2 \in \mathcal{C}_\mathbf{R}\}, \\ & && (14)\end{aligned}$$

where (a) is true because the signals $x_1(\hat{d}^{(j)})$, $\bar{y}_3^{(j)}$ in block j and the signals $x_2(\hat{d}^{(j)})$, $y_3^{(j+1)}$ in block $j+1$ are independent, (b) is due to that $p(x_1) = \frac{\mathbb{1}\{x_1 \in \mathcal{C}_S\}}{2^{NR_{DF}}}$ and $p(x_2) = \frac{\mathbb{1}\{x_2 \in \mathcal{C}_R\}}{2^{NR_{DF}}}$ being a uniform distribution over the elements of the source codebook $\mathcal{C}_S$ and the relay codebook $\mathcal{C}_R$ respectively and (c) follows that the decode-forward relay channel is i.i.d.. From the factorization (14), the joint decoding can be realized by connecting two decoders based on respective cost functions $p(\bar{y}_3|x_1)$ and $p(y_3|x_2)$ with iterative message passing between the information bits of two decoders.

The joint iterative belief propagation decoding model is proposed and shown in Fig. 5, where each decoder block corresponds to a factor graph. In this model, iterative message passing is applied not only within the factor graphs individually, but also between the factor graphs. The decoding is done in a multistage (progressive) manner. This is made possible by ensuring that the information about the source message at level k is contained in levels $\leq k$ at the destination. A sufficient condition is for the relay to encode the incoming message bits at each level into outgoing signals at the same level.

To avoid error bursts, we apply data interleavers π_k at the relay, whose inverse we denote with π_k^{-1}. Let $L_{l,i}^{S,k}$ and $R_{l,i}^{S,k}$ denote the right-to-left and left-to-right messages updated within factor graph for the source polar code, as well as $L_{l,i}^{R,k}$ and $R_{l,i}^{R,k}$ for the relay polar code, at level k. They are initialized as

$$\begin{aligned} L_{n+1,i}^{S,k} &\leftarrow \log\left(\frac{p(a_{k,i}=0|\bar{y}_{3,i}, \hat{a}_{1,i}, \cdots, \hat{a}_{k-1,i})}{p(a_{k,i}=1|\bar{y}_{3,i}, \hat{a}_{1,i}, \cdots, \hat{a}_{k-1,i})}\right), \\ L_{n+1,i}^{R,k} &\leftarrow \log\left(\frac{p(b_{k,i}=0|y_{3,i}, \hat{b}_{1,i}, \cdots, \hat{b}_{k-1,i})}{p(b_{k,i}=1|y_{3,i}, \hat{b}_{1,i}, \cdots, \hat{b}_{k-1,i})}\right), \\ R_{1,i}^{S,k},\ R_{1,i}^{R,k} &\leftarrow \begin{cases} \infty & i \in \mathcal{A}_k^c, \\ 0 & i \in \mathcal{A}_k. \end{cases}\end{aligned}$$

The remaining messages are initialized to zero. $e^{S,k}$ and $e^{R,k}$ denote the extrinsic messages for the source and the relay polar codes at level k, and are initialized to zero.

Let t_1 denote the maximum number of message passing iterations within factor graphs, and t_2 between factor graphs. The joint iterative belief propagation decoding at level k is described by the following steps:

1) Compute $R_{1,i}^{R,k} \leftarrow R_{1,i}^{R,k} + e_i^{R,k}$.
2) Update $L_{l,i}^{R,k}$ and $R_{l,i}^{R,k}$ via Eqs. (1)-(4) for t_1 iterations.
3) Compute $R_{1,i}^{R,k} \leftarrow R_{1,i}^{R,k} - e_i^{R,k}$ and the extrinsic messages $e_{\pi_k^{-1}(\ell)}^{S,k} \leftarrow L_{1,\ell}^{R,k}$, $\forall \ell \in \mathcal{A}_k$.
4) Compute $R_{1,i}^{S,k} \leftarrow R_{1,i}^{S,k} + e_i^{S,k}$
5) Update $L_{l,i}^{S,k}$ and $R_{l,i}^{S,k}$ via Eqs. (1)-(4) for t_1 iterations.
6) Compute $R_{1,i}^{S,k} \leftarrow R_{1,i}^{S,k} - e_i^{S,k}$ and the extrinsic messages $e_{\pi_k(\ell)}^{R,k} \leftarrow L_{1,\ell}^{S,k}$, $\forall \ell \in \mathcal{A}_k$.
7) If Steps 1-6 are repeated less than t_2 times, go to Step 1. Otherwise, stop to produce

$$\begin{aligned} \hat{u}_{k,i} &= \mathbb{1}_{\{i \in \mathcal{A}_k\} \cap \{R_{1,i}^{S,k} + L_{1,i}^{S,k} < 0\}}, \\ \hat{a}_{k,i} &= \mathbb{1}_{\{R_{n+1,i}^{S,k} + L_{n+1,i}^{S,k} < 0\}}, \\ \hat{b}_{k,i} &= \mathbb{1}_{\{R_{n+1,i}^{R,k} + L_{n+1,i}^{R,k} < 0\}}\end{aligned}$$

2) Successive Cancellation Decoding: When the source and the relay use the identical modulations, we have $x_1(d^{(j)}) = x_2(\tilde{d}^{(j)})$ with the assumption that $\tilde{d}^{(j)}$ is correctly decoded

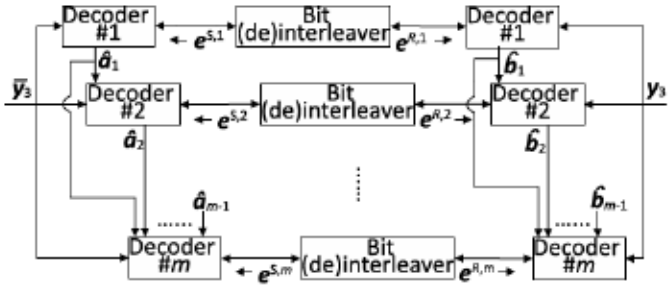


Fig. 5. Joint iterative belief propagation decoding model for the decode-forward relaying.

at the relay. In this case, the decode-forward relaying is considered to be a single-input double-output system model, where $x_1(d^{(j)})$ has two contaminated copies $\bar{y}_3^{(j)}$ and $y_3^{(j+1)}$. We combine $\bar{y}_3^{(j)}$ and $y_3^{(j+1)}$ according to the maximum ratio combining rule and let $y_{mrc}^{(j)}$ denote the result of combining. In this case, there is no interleaver at the relay. Using the knowledge of $\hat{u}_{k,\ell}^{(j)} = u_{k,\ell}, \forall \ell \in \mathcal{A}_k^c$, we recover $\{\hat{u}_{k,\ell}^{(j)} : \ell \in \mathcal{A}_k\}$ successively as

$$\hat{u}_{k,i} = \arg \max_{u \in \{0,1\}} p\big(u | y_{mrc}^{(j)}, \hat{a}_1^{(j)} \cdots \hat{a}_{k-1}^{(j)}, \hat{u}_{k,1}, \cdots, \hat{u}_{k,i-1}\big). \tag{15}$$

C. Dispersion Bound

In [6], the decoding word error probability P_e for the AWGN point-to-point channel as a function of the finite block length N and the code rate R was derived and given by

$$P_e = Q\left(\frac{C - R + O\left(\frac{\log_2(N)}{N}\right)}{\sqrt{\frac{F}{N}}}\right), \tag{16}$$

where the function $Q(\tau) = \int_\tau^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$. C and F denote the achievable rate and the dispersion of the channel. Denoting Z the unit-variance AWGN and $\{x_i\}$ the set of constellation points with size 2^m, we have [36]:

$$C = m - \frac{1}{2^m} \sum_{i=1}^{2^m} \mathbb{E}\left[\log_2\left(\sum_{j=1}^{2^m} e^{\|Z\|^2 - \|x_i + Z - x_j\|^2}\right)\right], \tag{17}$$

$$F = \frac{1}{2^m} \sum_{i=1}^{2^m} \mathrm{Var}\left[\log_2\left(\sum_{j=1}^{2^m} e^{\|Z\|^2 - \|x_i + Z - x_j\|^2}\right)\right]. \tag{18}$$

For a fixed code rate R_{DF} and block length N, we use Eqs. (16)-(18) to calculate the decoding error at the relay subject to a modified power constraint $g_{12}^2 \rho_s$. We also calculate the decoding error at the destination subject to the modified power constraint $g_{13}^2 \rho_s + \frac{g_{23}^2 \rho_r}{g_{12}^2 \rho_s + 1}$, which is obtained by applying maximum ratio combining on $\bar{y}_3^{(j)}$ and $y_3^{(j+1)}$. The overall word error probability can be derived from the errors at the relay and destination as follows:

$$\begin{aligned} P(E_D) &= P(E_D|E_R)\, P(E_R) + P(E_D|\bar{E}_R)\, P(\bar{E}_R) \\ &\approx P(E_R) + (1 - P(E_R)) P(E_D|\bar{E}_R) \end{aligned}$$

where E_R and E_D are block-error Bernoulli random variables at the relay and destination, respectively. We use $P(E_D|E_R) \approx 1$.

IV. Amplify-Forward

The relay helps the source by sending a scaled version of y_2 in the following block. We consider a linear relay with scaling coefficient $\frac{\sqrt{\rho_r}}{\sqrt{g_{12}^2 \rho_s + 1}}$, which guarantees that the average power constraint ρ_r is satisfied at the relay.

Define $\gamma \triangleq g_{23} \frac{\sqrt{\rho_r}}{\sqrt{g_{12}^2 \rho_s + 1}}$. In this paper, blocks are decoded in the sequence they are transmitted. In that case, $\hat{d}^{(j)}$ is recovered based on $\bar{y}_3(j)$ and $y_3^{(j+1)}$, where

$$\begin{aligned} \bar{y}_3^{(j)} &= y_3^{(j)} - \gamma g_{12} x_1(\hat{d}^{(j-1)}) \\ &= g_{13} x_1(d^{(j)}) + \gamma z_2 + z_3, \end{aligned} \tag{19}$$

$$y_3^{(j+1)} = \gamma g_{12} x_1(d^{(j)}) + g_{13} x_1(d^{(j+1)}) + \gamma z_2 + z_3. \tag{20}$$

It is also possible to do backward decoding (not used in the simulations of this paper). For completeness, we mention that in backward decoding, $\hat{d}^{(j)}$ would be recovered based on $y_3^{(j)}$ and $\tilde{y}_3^{(j+1)}$, where

$$y_3^{(j)} = g_{13} x_1(d^{(j)}) + \gamma g_{12} x_1(d^{(j-1)}) + \gamma z_2 + z_3. \tag{21}$$

$$\begin{aligned} \tilde{y}_3^{(j+1)} &= y_3^{(j+1)} - g_{13} x_1(\hat{d}^{(j+1)}) \\ &= \gamma g_{12} x_1(d^{(j)}) + \gamma z_2 + z_3. \end{aligned} \tag{22}$$

Since $x_1(d^{(j)})$ has two contaminated copies $\bar{y}_3^{(j)}$ and $y_3^{(j+1)}$, the amplify-forward relay channel is equivalent to a single-input double-output channel. Similar to Section III-B.2, by applying the maximum ratio combining on Eqs. (19), (20), a point-to-point AWGN channel is induced with signal to noise ratio (SNR) $\frac{g_{13}^2 \rho_s}{\gamma^2 + 1} + \frac{\gamma^2 g_{12}^2 \rho_s}{g_{13}^2 \rho_s + \gamma^2 + 1}$.

The source adopts multi-level polar-coded modulation in Section II-B to transmit the messages over the induced point-to-point channel. At the destination, using the knowledge of frozen bits, we recover the transmitted messages successively as shown in Eq. (15).

For a fixed code rate and block length N, we can use Eqs. (16)-(18) to calculate the amplify-forward dispersion bound at the destination by modifying the power constraint to $\frac{g_{13}^2 \rho_s}{\gamma^2 + 1} + \frac{\gamma^2 g_{12}^2 \rho_s}{g_{13}^2 \rho_s + \gamma^2 + 1}$.

V. Compress-Forward

In this section we present two different frameworks for multi-level compress-forward coded modulation. The first framework utilizes polarization for both error control and for Wyner-Ziv compression at the relay, which together permit successive decoding at the destination (in the sense of first decoding the relay signal, then the source signal). This framework is theoretically elegant, but practical design of Wyner-Ziv compression has produced uninspiring end-to-end performance for the compress-forward relay. The alternative is to allow joint decoding at the destination, which removes the necessity of Wyner-Ziv coding at the relay. In the second part of this section, we describe the joint decoding approach, which is the basis of our simulations.

A. Compress-Forward With Wyner-Ziv Compression

The relay uses Wyner-Ziv coding to compress y_2 into a binning index denoted λ and transmit $x_2(\lambda)$ in the following block. The achievable rate is given by [34]

$$R_{CF} < \max I(X_1; \hat{Y}_2, Y_3|X_2), \quad \text{subject to } I(Y_2; \hat{Y}_2|X_2, Y_3) \le I(X_2; Y_3), \quad (23)$$

where maximization is over $p(x_1)p(x_2)p(\hat{y}_2|x_2, y_2)$ and $\hat{Y}_2$ denotes the reconstructed version of Y_2.

The source and the relay employ multi-level coded modulation with m_1 and m_2 levels, respectively. The source and the relay transmit signals, X_1 and X_2, are related to the binary vectors $\boldsymbol{A} \triangleq [A_1, \ldots, A_{m_1}]$ and $\boldsymbol{B} \triangleq [B_1, \ldots, B_{m_2}]$, respectively, via bijective mappings. The achievable rate (23) can be expressed as

$$\begin{aligned} R_{CF} &\le \max \sum_{k_1=1}^{m_1} I(A_{k_1}; \hat{Y}_2, Y_3|\boldsymbol{B}, A_1, \cdots, A_{k_1-1}), \\ &\quad \text{subject to } I(Y_2; \hat{Y}_2|X_2, Y_3) \\ &\le \sum_{k_2=1}^{m_2} I(B_{k_2}; Y_3|B_1, \cdots, B_{k_2-1}), \end{aligned} \quad (24)$$

maximized over $\prod_{k_1=1}^{m_1} p(a_{k_1}) \prod_{k_2=1}^{m_2} p(b_{k_2}) p(\hat{y}_2|\boldsymbol{b}, y_2)$. Multi-level coding is optimal if

$$\prod_{k_1=1}^{m_1} p^*(a_{k_1}) \prod_{k_2=1}^{m_2} p^*(b_{k_2}) = p^*(\boldsymbol{a}) p^*(\boldsymbol{b}),$$

where $p^*(\cdot)$ represents the optimal distribution. The achievable rate at level k_1 denoted $R_{CF}(k_1)$ is less than or equal to $I(A_{k_1}; \hat{Y}_2, Y_3|\boldsymbol{B}, A_1, \cdots, A_{k_1-1})$.

1) Codebooks: At the source, polar codes constructed in Section II-B are adopted to transmit message over the point-to-point sub-channels denoted $W'_{k_1} : A_{k_1} \to \{\hat{Y}_2, Y_3, \boldsymbol{B}, A_1 \cdots A_{k_1-1}\}$. $\mathcal{D}_{k_1}$ denotes the corresponding information sets and $\frac{|\mathcal{D}_{k_1}|}{N} = R_{CF}(k_1)$. $U_{k_1,\ell} = u$, $\forall \ell \in \mathcal{D}^c_{k_1}$ are fixed and revealed to the destination.

The relay compresses $y_2^{(j)}$ into $\hat{y}_2^{(j)}$, conditioned on $x_2(\lambda^{(j-1)})$ which is known to both the relay and the destination. Since the destination has side information $y_3^{(j)}$ about $\hat{y}_2^{(j)}$, Wyner-Ziv coding is utilized to reduce the rate necessary to transmit $\hat{y}_2^{(j)}$. The combined effect is achieved by the polar code $\hat{y}_2 = sG_N$, where s and G_N denote a non-binary sequence and polar transform for non-binary alphabet [37, Chapter 3] respectively. Let $\mathcal{T}$ and $\mathcal{V}$ denote the information sets corresponding to $p(x_2 y_2|\hat{y}_2)$ and $p(x_2 y_3|\hat{y}_2)$ respectively. Due to the Markov chain $\hat{Y}_2 - (X_2, Y_2) - (X_2, Y_3)$, $\mathcal{V}$ is a subset of $\mathcal{T}$ and $\frac{|\mathcal{T}\backslash\mathcal{V}|}{N} = I(\hat{Y}_2; X_2, Y_2) - I(\hat{Y}_2; X_2, Y_3) = I(\hat{Y}_2; Y_2|X_2, Y_3)$. $\{S_\ell : \ell \in \mathcal{T}\backslash\mathcal{V}\}$ represents the compressed binning index. $S_\ell = s$, $\forall \ell \in \mathcal{T}^c$ are chosen uniformly at random over the set $[1 : |\mathcal{S}|]$, and known by the relay and the destination.

The relay also applies the polar codes constructed in Section II-B to transmit the compressed binning index over the point-to-point sub-channels denoted $W''_{k_2} : B_{k_2} \to \{Y_3, B_1 \cdots B_{k_2-1}\}$. $\mathcal{J}_{k_2}$ denotes the corresponding information sets where $\sum_{k_2=1}^{m_2} \frac{|\mathcal{J}_{k_2}|}{N} = I(Y_2; \hat{Y}_2|X_2, Y_3)$ and $\frac{|\mathcal{J}_{k_2}|}{N} \le I(W''_{k_2})$. $U_{k_2,\ell} = u$, $\forall \ell \in \mathcal{J}^c_{k2}$ are fixed and revealed to the destination.

2) Encoding and Decoding: In block j, the source demultiplexes the message $d^{(j)}$ into m_1 bit-streams according to $R_{CF}(k_1)$, uses $\{u^{(j)}_{k_1,\ell} : \ell \in \mathcal{D}_{k_1}\}$ to carry each bit-stream, and transmits $x_1(d^{(j)})$.

The relay first uses successive encoding [38] for the compression by setting $S^{(j)}_\ell = s_\ell$, $\forall \ell \in \mathcal{T}^c$ and choosing $S^{(j)}_i = s$, $\forall i \in \mathcal{T}$ randomly according to the probability $p(s|x_2(\lambda^{(j-1)}), \boldsymbol{y}_2^{(j)}, s_1, \cdots, s_{i-1})$. It then demultiplexes the binning index $\{s^{(j)}_\ell : \ell \in \mathcal{T}\backslash\mathcal{V}\}$, i.e. $\lambda^{(j)}$, into m_2 bit-streams according to $\frac{|\mathcal{J}_{k_2}|}{N}$, each carried by $\{u^{(j)}_{k_2,\ell} : \ell \in \mathcal{J}_{k_2}\}$. The relay transmits $x_2(\lambda^{(j)})$ in block $j+1$.

At the end of block $j+1$, knowing $\hat{u}^{(j)}_{k_2,\ell} = u_{k_2,\ell}$, $\forall \ell \in \mathcal{J}^c_{k_2}$, the destination first recovers $\{\hat{u}^{(j)}_{k_2,\ell} : \ell \in \mathcal{J}_{k_2}\}$ successively as

$$\hat{u}_{k_2,i} = \arg \max_{u \in \{0,1\}} p\big(u|\boldsymbol{y}_3^{(j+1)}, \hat{\boldsymbol{b}}_1^{(j)} \cdots \hat{\boldsymbol{b}}_{k_2-1}^{(j)}, \hat{u}_{k_2,1}, \cdots, \hat{u}_{k_2,i-1}\big).$$

Therefore, the estimated binning index $\hat{\lambda}^{(j)}$ (equivalently $\{\hat{s}^{(j)}_\ell : \ell \in \mathcal{T}\backslash\mathcal{V}\}$) is recovered. Using the knowledge of $\hat{s}^{(j)}_\ell = s_\ell$, $\forall \ell \in \mathcal{T}^c$ and the estimated binning index, it then recovers $\{\hat{s}^{(j)}_\ell : \ell \in \mathcal{V}\}$ successively as

$$\hat{s}_i = \arg \max_{s \in [1:|\mathcal{S}|]} p\big(s|\boldsymbol{x}_2(\hat{\lambda}^{(j-1)}), \boldsymbol{y}_3^{(j)}, \hat{s}_1, \cdots, \hat{s}_{i-1}\big).$$

Based on $\hat{\boldsymbol{y}}_2^{(j)} = \hat{\boldsymbol{s}}^{(j)} G_N$ and $\hat{u}^{(j)}_{k_1,\ell} = u_{k_1,\ell}$, $\forall \ell \in \mathcal{D}^c_{k_1}$, the destination recovers $\{\hat{u}^{(j)}_{k_1,\ell} : \ell \in \mathcal{D}_{k_1}\}$ successively as

$$\hat{u}_{k_1,i} = \arg \max_{u \in \{0,1\}} p\big(u|\hat{\boldsymbol{y}}_2^{(j)}, \boldsymbol{y}_3^{(j)}, \boldsymbol{x}_2(\hat{\lambda}^{(j-1)}), \hat{\boldsymbol{a}}_1^{(j)} \cdots \hat{\boldsymbol{a}}_{k_1-1}^{(j)}, \hat{u}_{k_1,1}, \cdots, \hat{u}_{k_1,i-1}\big).$$

B. Avoiding Wyner-Ziv Compression via Joint Decoding

To avoid the difficulties experienced by practical Wyner-Ziv implementations, we consider the strategy without Wyner-Ziv coding, where the relay quantizes y_2 and directly maps the quantized binning to a codeword $x_2(\lambda)$. It is known [39], [40], [41] that the same performance as Wyner-Ziv compression can be obtained as long as the destination performs joint decoding of the relay and source signals, instead of successive decoding.

For practical reasons, a scalar quantizer is used for compression. For scalar quantizer, a resolution of 2^{m_1} is adopted, since [42] showed that finer scalar quantization does not produce noticeable gain. As pointed out by [19], [43], optimizing the quantization alphabet according to the distortion metric may not be the best strategy for compress-forward. The reason is that the ultimate goal of the relay is to provide useful information for assisting the decoding at the destination, which may not be identical to producing maximally faithful local representations of y_2. In practice, we found that a scalar quantization alphabet that is a scaling of X_1 by a factor g_{12} produces the best results.

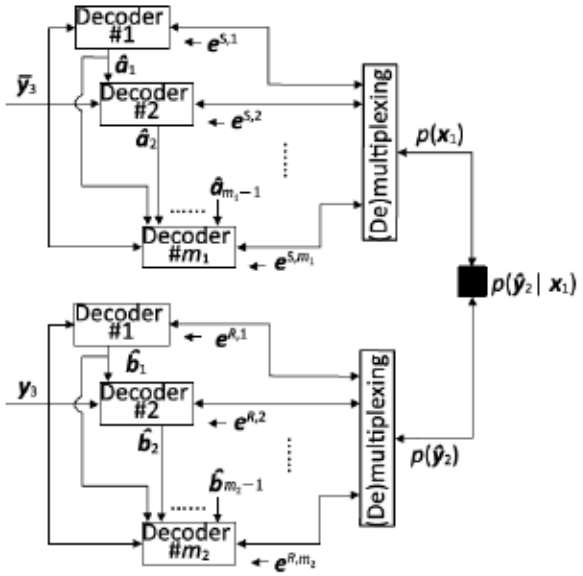


Fig. 6. Joint iterative belief propagation decoding model for the compress-forward relaying.

Joint Iterative Belief Propagation Decoding: At the end of block $j+1$, the destination jointly decodes the message $\hat{d}^{(j)}$ and the quantized binning $\hat{\lambda}^{(j)}$ based on $y_3^{(j+1)}$, the destination observation in block $j+1$, and $\bar{y}_3^{(j)}$, the destination observation in block j from which the relay transmission $x_2(\hat{\lambda}^{(j-1)})$ has been peeled off as shown in Eq. (10).

The destination searches for candidate codewords $\big(x_1(\hat{d}^{(j)}), x_2(\hat{\lambda}^{(j)})\big)$ that maximize the *a posteriori* probability $p\big(x_1(\hat{d}^{(j)})x_2(\hat{\lambda}^{(j)})|\bar{y}_3^{(j)}y_3^{(j+1)}\big)$ which can be factorized as

$$p(x_1x_2|\bar{y}_3y_3) = \frac{1}{p(\bar{y}_3)p(y_3)}\sum_{\hat{y}_2} p(x_1, x_2, \hat{y}_2, \bar{y}_3, y_3), \quad (25)$$

where

$$\begin{aligned} &p(x_1, x_2, \hat{y}_2, \bar{y}_3, y_3) \\ &\quad = p(\bar{y}_3, y_3|x_1, x_2, \hat{y}_2)p(x_1, x_2, \hat{y}_2), \\ &\quad \overset{(a)}{=} p(\bar{y}_3, y_3|x_1, x_2)p(x_1, x_2, \hat{y}_2), \quad (26) \\ &\quad = p(\bar{y}_3, y_3|x_1, x_2)p(x_2|x_1, \hat{y}_2)p(\hat{y}_2|x_1)p(x_1), \quad (27) \\ &\quad \overset{(b)}{=} p(\bar{y}_3, y_3|x_1, x_2)p(x_2|\hat{y}_2)p(\hat{y}_2|x_1)p(x_1), \quad (28) \\ &\quad \overset{(c)}{=} p(\bar{y}_3|x_1)p(y_3|x_2)p(x_2|\hat{y}_2)p(\hat{y}_2|x_1)p(x_1), \quad (29) \\ &\quad \overset{(d)}{\propto} p(\bar{y}_3|x_1)p(y_3|x_2)\mathbb{1}\{x_2 = \mathcal{C}_R(\hat{y}_2)\} \\ &p(\hat{y}_2|x_1)\mathbb{1}\{x_1 \in \mathcal{C}_S\}, \quad (30) \\ &\quad \overset{(e)}{=} \prod_{i=1}^{N} p(\bar{y}_{3,i}|x_{1,i})p(y_{3,i}|x_{2,i})\mathbb{1}\{x_2 = \mathcal{C}_R(\hat{y}_2)\} \\ &\quad p(\hat{y}_2|x_1)\mathbb{1}\{x_1 \in \mathcal{C}_S\}, \quad (31) \end{aligned}$$

where (a) follows the fact that the channel output only depends on its input, (b) is due to the Markov chain $\boldsymbol{X}_1 - \boldsymbol{Y}_2 - \boldsymbol{X}_2$, since in the present case x_2 is a function of $\hat{y}_2$, (c) follows that the signals transmitted in different blocks are independent (note that $x_2(\hat{\lambda}^{(j)})$ is in block $j+1$), (d) is due to $p(x_2|\hat{y}_2) = \mathbb{1}\{x_2 = \mathcal{C}_R(\hat{y}_2)\}$, i.e., the output of polar coding at the relay being a deterministic function $\mathcal{C}_R$ of its input, and $p(x_1) = \frac{\mathbb{1}\{x_1 \in \mathcal{C}_S\}}{2^{NR_{CF}}}$ being a uniform distribution over the elements of the source codebook $\mathcal{C}_S$, and (e) follows that the compress-forward relay channel is i.i.d..

Based on the factorization (31), the joint decoding can be realized by connecting two decoders whose cost functions are based on $p(\bar{y}_3|x_1)$ and $p(y_3|x_2)$ with iterative message passing between x_1 and $\hat{y}_2$ enabled by the term

$$p(\hat{y}_2|x_1) = \int\limits_{\hat{y}_2 = Q_a(y_2)} p(y_2|x_1)\, dy_2, \quad (32)$$

where Q_a represents the scalar quantizer. A joint iterative belief propagation decoding model is proposed and shown in Fig. 6, where each decoder block corresponds to a factor graph and the node characterized by $p(\hat{y}_2|x_1)$ connects the factor graphs for the source and the relay polar codes. The detailed message passing algorithm is described as follows.

Similar to Section III-B.1, let $L_{l,i}^{S,k_1}$, $R_{l,i}^{S,k_1}$ and $L_{l,i}^{R,k_2}$, $R_{l,i}^{R,k_2}$ denote the right-to-left and left-to-right message pairs updated within factor graphs for the source and the relay polar codes at level k_1 and k_2 respectively.

Unlike the multi-level coding in the decode-forward relaying, where the relay signal at each level only relates to the source signal at the same level, the scalar quantizer which is not in multi-level form couples the source and the relay signals at different levels. Therefore, during the stage of iterative massage passing between factor graphs, we use parallel independent decoding instead of multistage decoding. The messages are initialized as

$$\begin{aligned} L_{n+1,i}^{S,k_1} &\leftarrow \log\left(\frac{p(a_{k_1,i}=0|\bar{y}_{3,i})}{p(a_{k_1,i}=1|\bar{y}_{3,i})}\right), \\ L_{n+1,i}^{R,k_2} &\leftarrow \log\left(\frac{p(b_{k_2,i}=0|y_{3,i})}{p(b_{k_2,i}=1|y_{3,i})}\right), \\ R_{1,i}^{S,k_1} &\leftarrow \begin{cases} \infty & i \in \mathcal{D}_{k_1}^c, \\ 0 & i \in \mathcal{D}_{k_1}, \end{cases} \\ R_{1,i}^{R,k_2} &\leftarrow \begin{cases} \infty & i \in \mathcal{J}_{k_2}^c, \\ 0 & i \in \mathcal{J}_{k_2}. \end{cases} \end{aligned} \quad (33)$$

The remaining right-to-left and left-to-right messages are initialized to zero.

Let $p(x_1)$ and $p(\hat{y}_2)$ denote the soft information in the probability form for the signals x_1 and $\hat{y}_2$ respectively. Let e^{S,k_1} and e^{R,k_2} denote the extrinsic LLR messages passed to the factor graphs for the source and the relay at level k_1 and k_2 respectively, and initialize them to zero.

The iterative message passing between factor graphs is described by the following steps:

1) Compute $R_{1,i}^{R,k_2} \leftarrow R_{1,i}^{R,k_2} + e_i^{R,k_2}$ at all m_2 levels in parallel.
2) Update $L_{l,i}^{R,k_2}$ and $R_{l,i}^{R,k_2}$ via Eqs. (1)-(4) for t_1 iterations at all m_2 levels in parallel.
3) Compute $R_{1,i}^{R,k_2} \leftarrow R_{1,i}^{R,k_2} - e_i^{R,k_2}$ and produce $\{L_{1,\ell}^{R,k_2} : \ell \in \mathcal{J}_{k_2}\}$ at all m_2 levels in parallel.
4) Map $\{L_{1,\ell}^{R,k_2} : \ell \in \mathcal{J}_{k_2}\}$ from m_2 levels to $p(\hat{y}_2)$.

5) With $p(\hat{y}_2)$, compute $p(x_1)$ through the calculation of marginals of x_1 via Eq. (32).
6) Demultiplex $p(x_1)$ into m_1 extrinsic LLR streams e^{S,k_1} by computing $e_i^{S,k_1} \leftarrow log\big(\frac{p(a_{k_1,i}=0)}{p(a_{k_1,i}=1)}\big)$.
7) Compute $L_{n+1,i}^{S,k_1} \leftarrow L_{n+1,i}^{S,k_1} + e_i^{S,k_1}$ at all m_1 levels in parallel.
8) Update $L_{l,i}^{S,k_1}$ and $R_{l,i}^{S,k_1}$ via Eqs. (1)-(4) for t_1 iterations at all m_1 levels in parallel.
9) Compute $L_{n+1,i}^{S,k_1} \leftarrow L_{n+1,i}^{S,k_1} - e_i^{S,k_1}$ and produce $\{R_{n+1,\ell}^{S,k_1} : \ell = 1, \cdots, N\}$ at all m_1 levels in parallel.
10) Map $\{R_{n+1,\ell}^{S,k_1} : \ell = 1, \cdots, N\}$ from m_1 levels to $p(x_1)$.
11) With $p(x_1)$, compute $p(\hat{y}_2)$ through the calculation of marginals of $\hat{y}_2$ via Eq. (32).
12) Demultiplex $p(\hat{y}_2)$ into m_2 extrinsic LLR streams $\{e_\ell^{R,k_2} : \ell \in \mathcal{J}_{k_2}\}$ by computing $\log\big(\frac{p(u_{k_2}=0)}{p(u_{k_2}=1)}\big)$.
13) If Steps 1-12 are repeated less than t_2 times, go to Step 1. Otherwise, stop.

After iterative message passing via Steps 1-13, we use multistage decoding. We initialize $L_{n+1,i}^{S,k_1}$ and $L_{n+1,i}^{R,k_2}$ as follows:

$$\begin{aligned} L_{n+1,i}^{S,k_1} \leftarrow & \log\left(\frac{p(a_{k_1,i}=0|\bar{y}_{3,i}, \hat{a}_{1,i}, \cdots, \hat{a}_{k_1-1,i})}{p(a_{k_1,i}=1|\bar{y}_{3,i}, \hat{a}_{1,i}, \cdots, \hat{a}_{k_1-1,i})}\right) \\ & + \log\left(\frac{p(a_{k_1,i}=0|\hat{a}_{1,i}, \cdots, \hat{a}_{k_1-1,i})}{p(a_{k_1,i}=1|\hat{a}_{1,i}, \cdots, \hat{a}_{k_1-1,i})}\right), \\ L_{n+1,i}^{R,k_2} \leftarrow & \log\left(\frac{p(b_{k_2,i}=0|y_{3,i}, \hat{b}_{1,i}, \cdots, \hat{b}_{k_2-1,i})}{p(b_{k_2,i}=1|y_{3,i}, \hat{b}_{1,i}, \cdots, \hat{b}_{k_2-1,i})}\right), \end{aligned} \quad (34)$$

where the second term on the right hand side of Eq. (34) is based on $p(x_1)$ obtained in Step 5. We initialize $R_{1,i}^{S,k_1}$ according to Eq. (33). Further, we initialize $R_{1,i}^{R,k_2}$ to e_i^{R,k_2} obtained in Step 12. The remaining messages are initialized to zero.

We update $L_{l,i}^{S,k_1}$, $R_{l,i}^{S,k_1}$ and $L_{l,i}^{R,k_2}$, $R_{l,i}^{R,k_2}$ via Eqs. (1)-(4) for t_3 iterations. The decoder outputs are:

$$\begin{aligned} \hat{u}_{k_1,i} &= \mathbb{1}_{\{i \in \mathcal{D}_{k_1}\} \cap \{R_{1,i}^{S,k_1} + L_{1,i}^{S,k_1} < 0\}} \\ \hat{a}_{k_1,i} &= \mathbb{1}_{\{R_{n+1,i}^{S,k_1} + L_{n+1,i}^{S,k_1} < 0\}} \\ \hat{b}_{k_2,i} &= \mathbb{1}_{\{R_{n+1,i}^{R,k_2} + L_{n+1,i}^{R,k_2} < 0\}} \end{aligned}$$

In Section VII, we also show the error performance based on joint iterative belief propagation list decoding. The list decoder consists of L independent parallel joint iterative belief propagation decoders mentioned above. In each joint iterative belief propagation decoder, all decoder blocks in Fig. 6 are based on the same permuted factor graph. The output of joint iterative belief propagation decoder whose estimated codeword $x_1(\hat{d}^{(j)})$ is closest to $\bar{y}_3^{(j)}$ in Euclidean distance is selected as the output of list decoder. In our work, the strategy of selecting factor graph permutations is as follows: the original factor graph $h = [1, 2, \cdots, n]$ is always kept; the rest are selected from the permutations where the elements at both ends, i.e. 1 and n, are unchanged and the elements close to n are permuted in priority.

We apply the lower bound of the achievable rate in [34, Eq. 16.12] which shows that the modified power constraint is equal to $g_{13}^2\rho_s + \frac{g_{12}^2\rho_s g_{23}^2\rho_r}{g_{13}^2\rho_s + g_{12}^2\rho_s + g_{23}^2\rho_r + 1}$. We use Eqs. (16)-(18) for the calculation of the dispersion bound at the destination for a fixed code rate R_{CF} and block length N, with the modified power constraint mentioned above.

VI. Error Exponent Analysis

For the discussion and illuminating explanations of our simulation results in Section VII, a random coding error exponent is analyzed for multi-level coding. Let $d_k \in \mathcal{D}_k$, $1 \le k \le m$, denote the message sent on level k. The corresponding binary codeword at level k, denoted $a_k(d_k)$, has length N and is drawn from a codebook that is randomly generated, i.i.d., according to a Bernoulli random variable with parameter p_k. The code rate of level k is denoted by R_k, and the total rate of multi-level coding by $R = \sum_{k=1}^{m} R_k$.

In multistage decoding for the point-to-point channel, an error occurs if any of the messages $[d_1, \cdots, d_m]$ is decoded incorrectly. Let $\mathcal{E}_k$ denote the error event $d_k \neq \hat{d}_k$ on level k, and $\mathcal{E}_k^c$ denote the event that level k has been decoded correctly. Then, the error probability of multistage decoding is

$$\begin{aligned} P_e &\triangleq Pr\big(\cup_{k=1}^{m} \mathcal{E}_k\big) = Pr\big(\cup_{k=1}^{m} \big(\mathcal{E}_k \cap \mathcal{E}_{k-1}^c \cdots \cap \mathcal{E}_1^c\big)\big). \\ &\overset{(a)}{\le} \sum_{k=1}^{m} Pr\big(\mathcal{E}_k \cap \mathcal{E}_{k-1}^c \cdots \cap \mathcal{E}_1^c\big) \\ &= \sum_{k=1}^{m} \Big(Pr\big(\mathcal{E}_k | \cap_{l=1}^{k-1} \mathcal{E}_l^c\big) \cdot Pr\big(\cap_{l=1}^{k-1} \mathcal{E}_l^c\big)\Big), \\ &\le \sum_{k=1}^{m} Pr\big(\mathcal{E}_k | \cap_{l=1}^{k-1} \mathcal{E}_l^c\big), \end{aligned} \quad (35)$$

where (a) is due to the union bound and the terms

$$P_{ek} \triangleq Pr\big(\mathcal{E}_k | \cap_{l=1}^{k-1} \mathcal{E}_l^c\big) \quad (36)$$

inside the summation on the right-hand side of Eq. (35) are the probability of error on level k conditioned on all previous levels being decoded correctly.

We now import an error exponent analysis [44] developed by Gallager motivated by successive cancellation decoding in the multiple access channel (MAC). In [44], the exponent of the error probability of one message in MAC was calculated while (one or more) other messages are assumed known or correctly decoded at the receiver. This scenario also describes the error probability in Eq. (36), therefore we import this error exponent for our purposes and denote it $E^{(k)}(R_k)$. According to [44],

$$P_{ek} \le 2^{-N \cdot \left(E_0^{(k)}(\rho) - \rho R_k\right)}, \quad (37)$$

where the parameter $E_0^{(k)}(\rho)$ was calculated in [44] as follows:

$$\begin{aligned} E_0^{(k)}(\rho) \triangleq & -\log_2 \Bigg[\sum_{a_1, \cdots, a_{k-1}} \prod_{l=1}^{k-1} p_l \int_y \\ & \times \Bigg(\sum_{a_k} p_k \cdot p\big(y | a_k, a_1, \cdots, a_{k-1}\big)^{\frac{1}{1+\rho}} \Bigg)^{1+\rho} \mathrm{d}y \Bigg]. \end{aligned} \quad (38)$$

Eq. (37) is true for any ρ, $0 \le \rho \le 1$ and thus

$$P_{ek} \le 2^{-N \cdot \max\limits_{0\le\rho\le1}\left(E_0^{(k)}(\rho)-\rho R_k\right)}, \tag{39}$$

which leads to Gallager's random coding error exponent for level k

$$E^{(k)}(R_k) \triangleq \max_{0\le\rho\le1}\left(E_0^{(k)}(\rho) - \rho R_k\right). \tag{40}$$

By using Eqs. (39) and (40), we can further bound Eq. (35) as follows

$$P_e \le \sum_{k=1}^{m} P_{ek} \le \sum_{k=1}^{m} 2^{-N\cdot E^{(k)}(R_k)} \le m \cdot 2^{-N\cdot \min\limits_k E^{(k)}(R_k)}. \tag{41}$$

This result was developed for a random code drawn according to an arbitrary distribution $p_1, \ldots, p_m$ (see Eq. (38)). Therefore, the exponent of error in Eq. (41) can be tightened via:

$$\max_{p_1,\cdots,p_m} \min_k E^{(k)}(R_k). \tag{42}$$

The above bound holds for arbitrary selection of rates $[R_1, \cdots, R_m]$. Therefore, constrained on the total rate R, the exponent in Eq. (42) can be further tightened by proper rate allocation, which we show below and denote $E^*(R)$

$$E^*(R) \triangleq \max_{R_1,\cdots,R_m} \max_{p_1,\cdots,p_m} \min_k E^{(k)}(R_k), \quad \text{subject to } \sum_K R_k = R \tag{43}$$

Thus, we conclude that there exists a random ensemble of codes for which the overall error probability of multistage decoding is bounded by

$$P_e \le m \cdot 2^{-N\cdot E^*(R)} = 2^{-N\cdot(E^*(R)-o(1))}, \tag{44}$$

and therefore, this is a lower bound on the error exponent for multi-level coding with multistage decoding.[5]

In the decode-forward relaying, we need to consider the decoding error events at the relay and the destination. The Gallager's random coding error exponent of the decode-forward relay channel is $E^*_{DF}(R_{DF}) = \min\{E^*_R(R_{DF}), E^*_D(R_{DF})\}$, where $E^*_R(R_{DF})$ and $E^*_D(R_{DF})$ denote the errror exponents corresponding to the decoding error events at the relay and the destination. The calculations of $E^*_R(R_{DF})$ and $E^*_D(R_{DF})$ via Eq. (43) or the calculations of $E_R^{(k)}\big(R_{DF}(k)\big)$ and $E_D^{(k)}\big(R_{DF}(k)\big)$ at level k via Eq. (40) are subject to respective power constraints $g_{12}^2\rho_s$ and $g_{13}^2\rho_s + \frac{g_{23}^2\rho_r}{g_{13}^2\rho_s+1}$.

The error exponent analysis for the amplify-forward relaying is omitted, since the amplify-forward relay channel is equivalent to a point-to-point channel.

For the compress-forward relaying, we only decode at the destination and similarly under multi-level coding with multistage decoding, the error exponent at level k_1 is evaluated via Eq. (40) with the parameter

$$E_0^{(k)}(\rho) = -\log_2\left\{\mathbb{E}\left[\sum_{\hat{y}_2\in\hat{\mathcal{Y}}_2}\int_{Y_3}\left(\sum_{a_{k_1}} p_{k_1} \times p\big(\hat{y}_2, y_3 | a_{k_1}, a_1, \cdots, a_{k_1-1}, x_2\big)^{\frac{1}{1+\rho}}\right)^{1+\rho} \mathrm{d}y_3\right]\right\}, \tag{45}$$

where the expectation is over x_2 and $[a_1, \cdots, a_{k_1-1}]$.

[5]The developments from Eq. (41) to (44) closely follow [45], and are included here for clarity and completeness.

VII. Simulations

A. Nomenclature

We use the following acronyms throughout the captions and legends in this section.

PAC	Polarization-adjusted convolutional code
MLC	Multi-level coding
BP	Belief propagation decoding
BPL-32	Belief propagation, size-32 list decoding
SP	Ungerboeck's set partitioning
Gray	Gray labeling
SC	Successive cancellation decoding
SCL-32	Successive cancellation, size-32 list decoding
HD	Half-duplex
FD	Full-duplex
PID	Parallel independent decoding

B. Results

The expectation and variance required in Eqs. (17) and (18) for computing the dispersion bounds are calculated by using Monte Carlo simulation. The relay channel consists of two transmitters and two receivers, therefore might have two different SNRs. To produce a traditional error plot, the transmit powers, channel gains, and noise variances must be tied to produce one SNR variable. In each of our experiments, the channel coefficients g_{13}, g_{12} and g_{23} are constants (to be specified for each experiment), the noise powers at the relay and destination are one, and the transmit powers at the source and the relay are identical. The SNR variable in the experiments is controlled by sweeping the source/relay power across the range that produces the prescribed SNR. However, since the rates of the codebooks at the source and the relay are not necessarily the same, the E_b/N_0 scale in the plots refers to the source codebook and the source-relay channel. This is further emphasized by the label in each plot.

The construction of the polar codes at all levels follows the ordered sequence in 5G standard [14]. In our simulations, the blocks are decoded in the order they were transmitted. It is also possible to employ backward decoding, but this would add to the overall decoding delay, and is not considered in this paper. The block Markov coding has $b = 20$ transmission blocks. The number of iterations in joint iterative belief propagation decoding are set to be $t_1 = 1$, $t_2 = 20$ and $t_3 = 20$.

Fig. 7 presents the BER of the proposed multi-level polar-coded modulation under decode-forward. We have 16-QAM at the source and the relay, the sum rate $R_{DF} = 2$ bits/s/Hz, block length 512 and $g_{13} = g_{12} = g_{23} = 1$. The rate allocations via chain rule have rates $0.62/0.38/0.62/0.38$ bits/s/Hz and $0.04/0.35/0.65/0.96$ bits/s/Hz for Gray labeling and Ungerboeck's set partitioning, respectively.

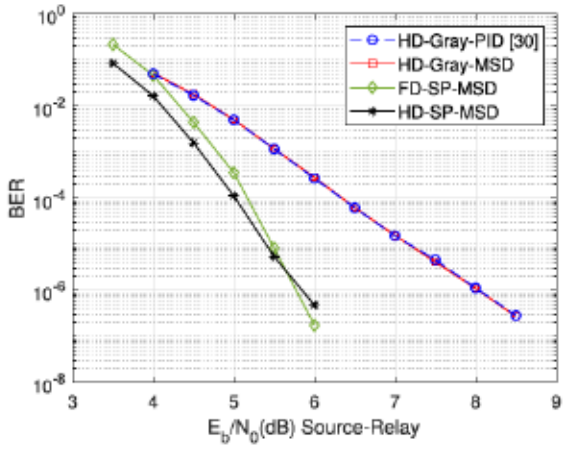


Fig. 7. Decode-forward, 16-QAM, 2 bits/s/Hz, block length 512, successive cancellation decoding.

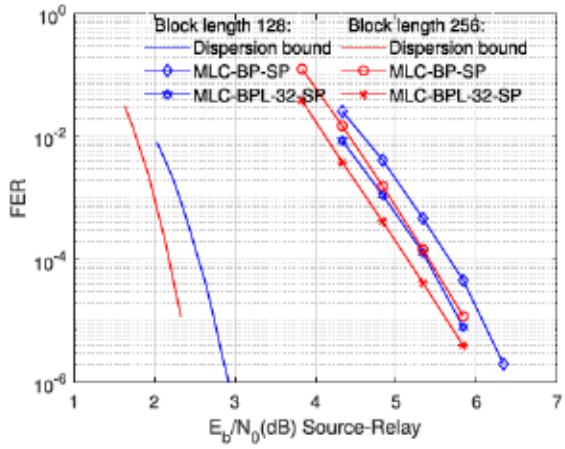


Fig. 10. Full-duplex compress-forward, 3.4 bits/s/Hz, 16-QAM at the source, 32-QAM at the relay.

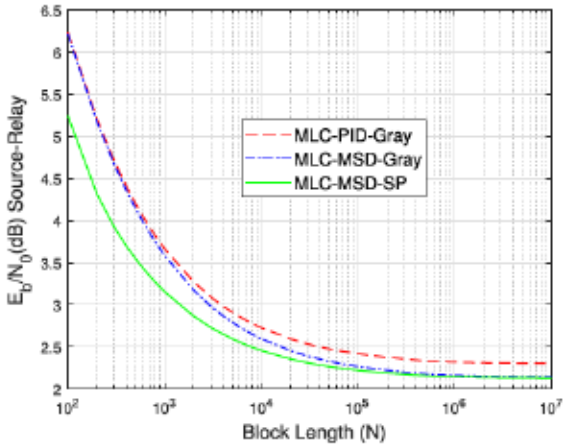


Fig. 8. Half-duplex DF, FER=10^{-3}, 2 bits/s/Hz.

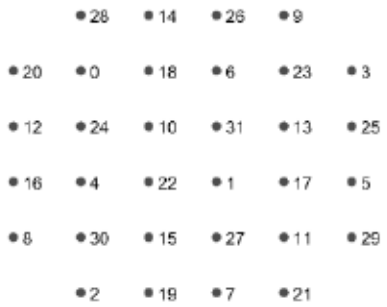


Fig. 11. 32-QAM, Ungerboeck's set partitioning.

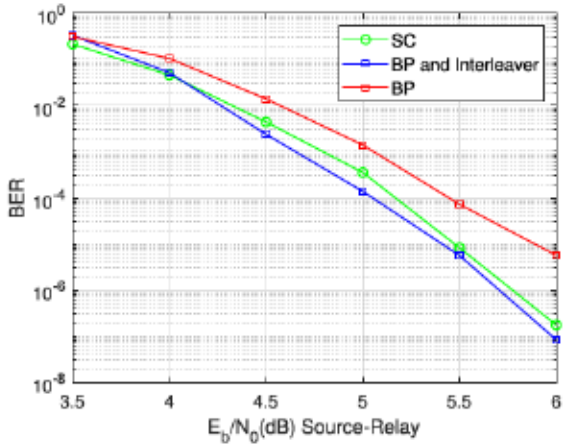
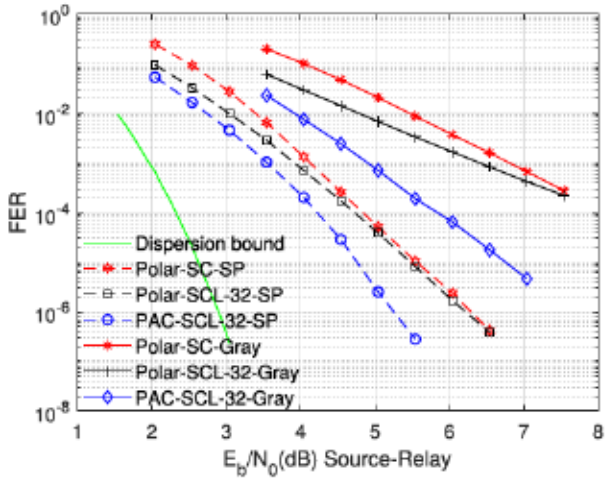


Fig. 9. Full-duplex DF, $N = 512$, 16-QAM, 2 bits/s/Hz, Ungerboeck's set partitioning.

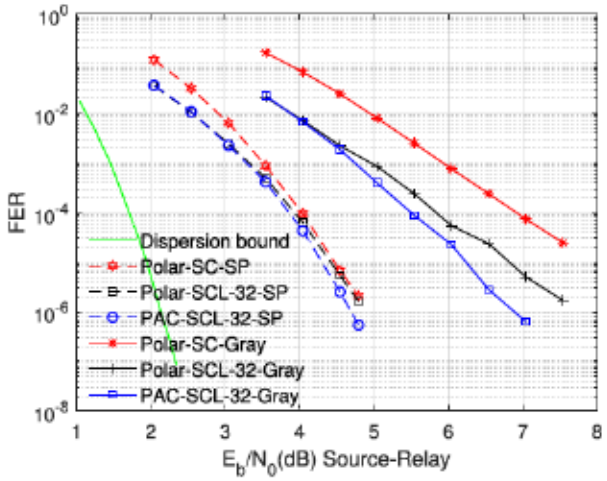
The closest work in the literature to the results of this paper appeared in [30], which is in orthogonal half-duplex mode with Gray labeling and parallel independent decoding. Compared with [30], our half-duplex result (see Fig. 7) shows the same performance by using gray labeling and as much as 2.5dB gain by using Ungerboeck's set partitioning. However, according to Eq. (5), multi-level coding with multistage decoding always preserves the capacity, regardless of the labeling of constellation points, which means the red and the black BER curves should have been very close, but they are not.

Fig. 7 shows a difference in the multistage decoding performance under Gray labeling and set-partitioning that is not visible through the multi-level capacity analysis via the chain rule. To shed further light on these differences, we utilize the random coding error exponents calculated earlier for the decode-forward relaying. This also provides an opportunity for another comparison against parallel independent decoding [30] across different block lengths. In Fig. 8, using $R_{DF} = 2$ bits/s/Hz and word error probability 10^{-3}, we compute and plot the required E_b/N_0 arising out of the random coding bounds. The comparison shows the relative performance of multistage decoding under Gray labeling and set-partitioning, confirming the BER simulations. The random coding bounds on the error of multistage decoding and parallel independent decoding also agrees with the BER simulations.

In Fig. 7, our full-duplex and half-duplex BER curves cross, which is deserving of an explanation. At low SNR, our full-duplex method performs worse than our half-duplex method; this is due to error propagation across message blocks in block Markov coding, a technique that is needed for full duplex relaying but is not present in half-duplex. At higher

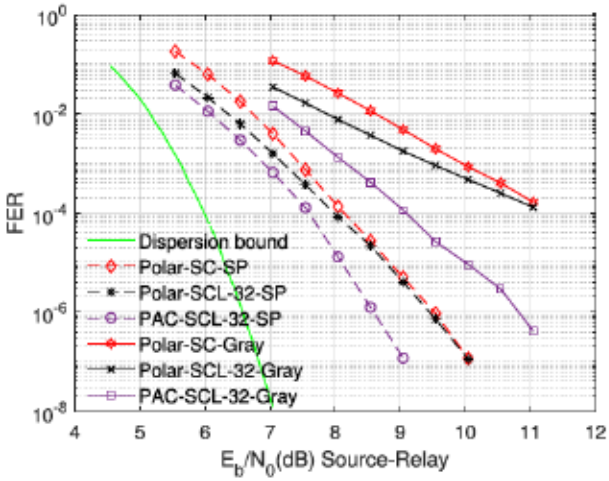


(a) Block length 128

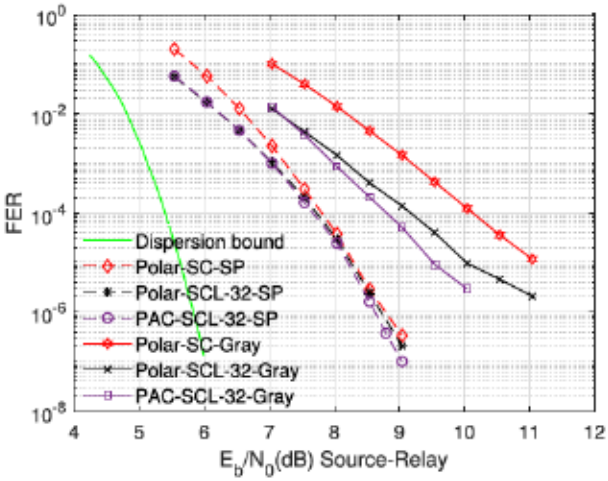


(b) Block length 256

Fig. 12. Full-duplex decode-forward, 8-PSK, 0.98 bits/s/Hz.



(a) Block length 128



(b) Block length 256

Fig. 13. Full-duplex amplify-forward, 8-PSK, 0.98 bits/s/Hz.

E_b/N_0, error propagation is less prominent, so the full-duplex mode starts to outperform the half-duplex mode.

Fig. 9 compares the BER of successive cancellation decoding and joint iterative belief propagation decoding with/without interleaver under decode-forward. It shows that belief propagation decoding with interleaver at the relay slightly outperforms successive cancellation decoding, while there is a 0.5dB loss at FER 10^{-5} without using the interleaver. Therefore, the interleaver is indispensable for joint iterative belief propagation decoding in order to compete with successive cancellation decoding.

Fig. 10 shows the frame error rate (FER) of the full-duplex compress-forward relaying with $g_{13} = 1$, $g_{12} = 0.72$ and $g_{23} = 11$, at block lengths 128 and 256. The source and the relay respectively use 16-QAM and 32-QAM under Ungerboeck's set partitioning, as well as have the sum rates $R_{CF} = 3.4$ bits/s/Hz with rate allocation $0.51/0.90/0.995/0.995$ bits/s/Hz and 4 bits/s/Hz with rate allocation $0.57/0.83/0.86/0.87/0.87$ bits/s/Hz. The constellation points for 32-QAM following Ungerboeck's set partitioning is shown in Fig. 11.

Previous work in this area using LDPC codes at block length 64,800 [19] has shown that the shaping gain given up by scalar quantization can be captured by using trellis coded quantization with Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm. However, in the short-block length regime, trellis coded quantization produced disappointing results according to our experiments.

C. Error Performance With PAC Codes

The PAC code is constructed by the concatenation of a rate-1 convolutional code and a polar code. The encoding begins by mapping u via the convolutional encoder to an output word which is then mapped to a codeword x via the polar transform $G_2^{\otimes n}$. The Reed-Muller rule is used to determine the information set and the frozen set. The frozen bits are known by the decoder. Sequential decoding [7], successive list

decoding [46], [47] and Viterbi decoding [48] were developed for PAC codes. Our work applies successive list decoding which follows the rules of regular successive list decoding except that extra shift registers are required to support the convolutional re-transform.

The performance of PAC codes for the full-duplex decode-forward and amplify-forward relaying is shown in Figs. 12 and 13. The Reed-Muller rule [7] and the convolutional code characterized by the impulse response $(1, 0, 1, 1, 0, 1, 1)$ are utilized to construct PAC codes at all levels. In these simulations, channel coefficients are $g_{13} = 1$, $g_{12} = 4$, $g_{23} = 4$, and the block length is 128 in Figs. 12b and 13b, and 256 in Figs. 12b and 13b. The source and the relay transmit using 8-PSK at the sum rate 0.98 bits/s/Hz with rate allocations $0.397/0.398/0.185$ bits/s/Hz for Gray labeling and $0.01/0.25/0.72$ bits/s/Hz for Ungerboeck's set partitioning. PAC codes demonstrate better performance than polar codes under both decode-forward and amplify-forward. In Figs. 12a and 13a, PAC codes have approximately 2dB gain for Gray labeling at FER 10^{-4} and 1dB gain for Ungerboeck's set partitioning at FER 10^{-6}. In Figs. 12b and 13b, PAC codes have approximately 1dB gain for Gray labeling at FER 10^{-6}.

Joint iterative belief propagation decoding for the compress-forward relaying using PAC codes did not produce promising results in our experiments.

VIII. Conclusion

In this paper, we propose short-block length multi-level polar-coded modulation for the decode-forward, amplify-forward and compress-forward relaying. Under decode-forward, we use multi-level polar-coded modulation at the source and the relay. We transmit the recovered message instead of its binning index at the relay and develop two decoding algorithms to jointly recover the source and the relay codewords at the destination. We develop multi-level polar-coded modulation with PAC code and successive list decoding to improve the performance. The simulation results show our method has a significant performance improvement over prior work, and the error exponent analysis explains the improvement. For amplify-forward relaying, we develop multi-level polar-coded modulation with PAC code at the source and successive list decoding at the destination. Under compress-forward, we use multi-level polar-coded modulation at the source and the relay. We use scalar quantization instead of Wyner-Ziv coding, and develop joint decoding to recover the source message and quantized bits. Joint decoding is realized by parallel independent decoding to produce extrinsic messages, followed by multistage decoding to recover the estimated codewords. The dispersion bounds for the full-duplex decode-forward, amplify-forward and compress-forward relaying are also presented.

References

[1] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Aug. 2016.

[2] T. Koike-Akino, Y. Wang, S. C. Draper, K. Sugihara, and W. Matsumoto, "Bit-interleaved polar-coded OFDM for low-latency M2M wireless communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.

[3] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *Proc. IEEE 1st Int. Conf. 5G Ubiquitous Connectivity*, Nov. 2014, pp. 146–151.

[4] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[5] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.

[6] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[7] E. Arıkan, "From sequential decoding to channel polarization and back again," 2019, *arXiv:1908.09594*.

[8] E. Arıkan, "Polar codes : A pipelined implementation," in *Proc. 4th ISBC*, vol. 2010, pp. 11–14.

[9] A. Elkelesh, M. Ebada, S. Cammerer, and S. T. Brink, "Belief propagation list decoding of polar codes," *IEEE Commun. Lett.*, vol. 22, no. 8, pp. 1536–1539, Aug. 2018.

[10] M. Geiselhart, A. Elkelesh, M. Ebada, S. Cammerer, and S. ten Brink, "CRC-aided belief propagation list decoding of polar codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 395–400.

[11] M. Seidl, A. Schenk, C. Stierstorfer, and J. B. Huber, "Polar-coded modulation," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4108–4119, Oct. 2013.

[12] E. Arıkan, "On the origin of polar coding," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 2, pp. 209–223, Feb. 2016.

[13] J. Dai, J. Piao, and K. Niu, "Progressive rate-filling: A framework for agile construction of multilevel polar-coded modulation," *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 1123–1127, Feb. 2021.

[14] *Multiplexing and Channel Coding*, Standard TS 38.212, ETSI, Release 15, 3GPP, Dec. 2019.

[15] C. I. Ionita, M. Mansour, J. C. Roh, and S. Hosur, "On the design of binary polar codes for high-order modulation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 3507–3512.

[16] Q. Zhang, A. Liu, X. Pan, and Y. Zhang, "Symbol-based belief propagation decoder for multilevel polar coded modulation," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 24–27, Jan. 2017.

[17] Y. Yajima and H. Ochiai, "On design of multilevel coded modulation based on CRC-concatenated polar codes," in *Proc. IEEE 17th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2020, pp. 1–6.

[18] L. Jiménez Rodríguez, N. H. Tran, and T. Le-Ngoc, "Performance of full-duplex AF relaying in the presence of residual self-interference," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1752–1764, Sep. 2014.

[19] H. Wan, A. Host-Madsen, and A. Nosratinia, "Compress- and-forward via multilevel coding and trellis coded quantization," *IEEE Commun. Lett.*, vol. 25, no. 4, pp. 1163–1167, Apr. 2021.

[20] A. A. Abotabl and A. Nosratinia, "Multilevel coded modulation for the full-duplex relay channel," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3543–3555, Jun. 2018.

[21] S. S. Ullah, S. C. Liew, G. Liva, and T. Wang, "Short-packet physical-layer network coding," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 737–751, Feb. 2020.

[22] Z. Chen, B. Xia, Z. Hu, and H. Liu, "Design and analysis of multi-level physical-layer network coding for Gaussian two-way relay channels," *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 1803–1817, Jun. 2014.

[23] M. H. Azmi, J. Yuan, G. Lechner, and L. K. Rasmussen, "Design of multi-edge-type bilayer-expurgated LDPC codes for decode- and-forward in relay channels," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 2993–3006, Nov. 2011.

[24] H. Khodaiemehr, D. Kiani, and M.-R. Sadeghi, "LDPC lattice codes for full-duplex relay channels," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 536–548, Feb. 2017.

[25] S. Takabe, T. Wadayama, and M. Hayashi, "Asymptotic behavior of spatial coupling LDPC coding for compute- and-forward two-way relaying," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4063–4072, Jul. 2020.

[26] T. Van Nguyen, A. Nosratinia, and D. Divsalar, "Bilayer protograph codes for half-duplex relay channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1969–1977, May 2011.

[27] R. Blasco-Serrano, R. Thobaben, M. Andersson, V. Rathi, and M. Skoglund, "Polar codes for cooperative relaying," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3263–3273, Feb. 2015.

[28] D. S. Karas, K. N. Pappi, and G. K. Karagiannidis, "Smart decode- and-forward relaying with polar codes," *IEEE Commun. Lett.*, vol. 3, no. 1, pp. 62–65, Feb. 2014.
[29] N. Madhusudhanan and L. Nithyanandan, "Compress- and-forward relaying with polar codes for LTE—A system," in *Proc. Int. Conf. Commun. Signal Process.*, Apr. 2014, pp. 798–802.
[30] X. Ma, L. Li, M. Zhu, W. Chen, and Z. Han, "Multilevel polar-coded modulation based on cooperative relaying," in *Proc. 9th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2017, pp. 1–4.
[31] A. Ingber and M. Feder, "On the optimality of multilevel coding and multistage decoding," in *Proc. IEEE 25th Conv. Electr. Electron. Engineers Isr.*, Dec. 2008, pp. 731–735.
[32] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Commun. Lett.*, vol. 13, no. 7, pp. 519–521, Jul. 2009.
[33] P. Trifonov, "Efficient design and decoding of polar codes," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3221–3227, Aug. 2012.
[34] A. El Gamal and Y. H. Kim, *Network Information Theory*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[35] L. Wang, "Polar coding for relay channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 1532–1536.
[36] E. C. Song and G. Yue, "Finite blocklength analysis for coded modulation with applications to link adaptation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–7.
[37] E. Sasoglu, "Polar coding theorems for discrete systems," Ph.D. dissertation, Ecole Polytechnique Federale de Lausanne, Vaud, Switzerland, Nov. 2011.
[38] S. B. Korada and R. L. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1751–1768, Apr. 2010.
[39] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.
[40] G. Kramer and J. Hou, "Short-message quantize-forward network coding," in *Proc. 8th Int. Workshop Multi-Carrier Syst. Solutions (MC-SS)*, Herrsching, Germany, May 2011, pp. 1–3.
[41] P. Zhong and M. Vu, "Compress-forward without Wyner-Ziv binning for the one-way and two-way relay channels," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Monticello, IL, USA, Sep. 2011, pp. 426–433.
[42] J. Amjad, M. Uppal, and S. Qaisar, "Multi-level compress and forward coding for half-duplex relays," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Anaheim, CA, USA, Dec. 2012, pp. 4536–4541.
[43] A. Chakrabarti, A. Sabharwal, and B. Aazhang, "Practical quantizer design for half-duplex estimate- and-forward relaying," *IEEE Trans. Commun.*, vol. 59, no. 1, pp. 74–83, Jan. 2011.
[44] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 2, pp. 124–142, Mar. 1985.
[45] A. Ingber and M. Feder, "Capacity and error exponent analysis of multilevel coding with multistage decoding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Sep. 2009, pp. 1799–1803.
[46] H. Yao, A. Fazeli, and A. Vardy, "List decoding of Arikan's PAC codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Oct. 2020, pp. 443–448.
[47] M. Rowshan, A. Burg, and E. Viterbo, "Polarization-adjusted convolutional (PAC) codes: Sequential decoding vs list decoding," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1434–1447, Feb. 2021.
[48] M. Rowshan and E. Viterbo, "List Viterbi decoding of PAC codes," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2428–2435, Mar. 2021.

Heping Wan (Member, IEEE) received the B.S. and M.S. degrees in telecommunication engineering from Xidian University, Shaanxi, China, and the Ph.D. degree in electrical engineering from The University of Texas at Dallas, Richardson, TX, USA. His research interests include information theory, coding theory, and their applications.

Aria Nosratinia (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana–Champaign in 1996. He is currently an Erik Jonsson Distinguished Professor and an Associate Head of the Electrical Engineering Department, The University of Texas at Dallas. He has held visiting appointments at Princeton University, Rice University, and UCLA. His research interests lie in the broad area of information theory and signal processing, with applications in wireless communications, data security and privacy. He is a fellow of IEEE for contributions to multimedia and wireless communications. He has served as an Editor and an Area Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and an Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE SIGNAL PROCESSING LETTERS, *IEEE Wireless Communications* (Magazine), and *Journal of Circuits, Systems, and Computers*. He has received the National Science Foundation Career Award and the Outstanding Service Award from the IEEE Signal Processing Society, Dallas Chapter. He has served as the Secretary for the IEEE Information Theory Society, Treasurer for ISIT, the Publications Chair for the IEEE Signal Processing Workshop, and a member of the technical committee for a number of conferences. He was the General Co-Chair of IEEE Information Theory Workshop in 2018. He is a Registered Professional Engineer in the state of Texas and a Clarivate Analytics Highly Cited Researcher.