Bit-Interleaved Polar Coded Modulation with Iterative Decoding

Souradip Saha, Matthias Tschauner, Marc Adrat
Fraunhofer FKIE
Wachtberg 53343, Germany
Email: *firstname.lastname*@fkie.fraunhofer.de

Tim Schmitz, Peter Jax, Peter Vary
Institute of Communication Systems, RWTH Aachen University
Aachen 52056, Germany
Email: *lastname*@iks.rwth-aachen.de

Abstract—**Polar Codes are a recently proposed class of linear block error correction codes. They are provably capacity achieving codes over Binary Discrete Memoryless Channels (B-DMC) and have hence garnered a lot of interest from the scientific community. It is also a proposed channel coding method for 5G technology. Bit-Interleaved Coded Modulation with Iterative Decoding (BICM-ID) is a well known design to improve the error correcting performance of underlying channel codes over continuous channels especially Additive White Gaussian Noise (AWGN) channels. The novel idea in this paper, is to combine these powerful error correcting techniques i.e. integrate Polar Codes in a BICM-ID design to produce a high performance Bit-Interleaved Polar Coded Modulation with Iterative Decoding (BIPCM-ID) system. The error correcting performance of such a BIPCM-ID system has been analyzed through simulations over AWGN channel and multiple modulation schemes. Additionally error floor removal has been implemented and system performance has been discussed.**

I. INTRODUCTION

Arikan in his paper [1], proposed a new channel coding scheme by "polarizing" channels w.r.t. their capacities. The process of channel polarization helps to classify every bit channel w.r.t. channel capacity which is used to create a scheme of encoding information bits into a codeword. This form of classification provides a new family of linear block codes called Polar Codes.

BICM-ID is a state-of-the-art error correcting scheme. The coded-modulation scheme was proposed in [2] for improving error correction capabilities by using channel coding and modulation schemes as a combined unit instead of independent modules. Adding diversity to the codes by interleaving, should improve the error correction performance as proposed in [3], to hence provide the Bit Interleaved Coded Modulation (BICM) design described in [4]. BICM's demerits can be overcome by exchanging extrinsic information between multiple modules of the Iterative Decoding (ID) chain iteratively, to improve error correcting capability over every iteration [5]. Such a BICM-ID scheme is helpful for error correction of codes received over AWGN and Fading channels.

The idea is to encode the input data bits into codewords generated by polar encoding, interleaving and then modulating them at the transmitter end. Similarly, at the receiver end polar decoder, demodulator, interleaver and de-interleaver iteratively perform their tasks and exchange extrinsic information to produce codeword estimates. The aim in this paper is to develop such a novel BIPCM-ID system and analyze its error correcting capability.

Although Arikan's original idea of channel polarization is limited only to Binary Erasure Channels (BEC) the BIPCM-ID needs to be analyzed over AWGN channels. Thus, a modified approach of channel polarization for AWGN channels is used. The polar decoder which is to be integrated in an ID chain should be a Soft-Input Soft-Output (SISO) decoder so that it can produce extrinsic information to be exchanged with, as well as process the inputs from, the other modules in the ID chain. Modifications can also be made to remove the error floor altogether. These issues have been addressed in this paper.

The rest of the paper is organized as follows. In Section II, an overview of Polar Codes and BICM-ID is provided for the reader to get a basic understanding of these concepts. Section III describes the novel approach of integrating Polar Codes in a BICM-ID design by generating the extrinsic information from the decoder module to be used for ID and how the channels are polarized for a AWGN channel. Section IV, provides the simulation results of the BIPCM-ID system, with detailed discussions about the parameters used and the corresponding performance results obtained. Section V provides the key points which merit further research and the existing limitations which can be improved upon. The concluding remarks are mentioned in Section VI.

Notations: Letters in bold fonts denote vectors. X and Y denote the Random Variables (RV) corresponding to the input and output of a channel respectively. W denotes a channel as well as the Probability Density Functions (PDF) of the RVs across the channel. N denotes the length of a codeword. Z denotes the Bhattacharyya parameter. L denotes the Log-Likelihood Ratio (LLR) values.

II. PRELIMINARIES

In this Section, a general overview of the concepts of Polar Codes and BICM-ID are provided which lay the foundation for developing the target BIPCM-ID system.

A. Polar Codes

1) Channel Polarization: Channel polarization is a technique used to segregate multiple channels with identical capacities, such that the channels are polarized w.r.t. their capacities, i.e. every channel can be categorized as a high capacity or a low

capacity channel after channel polarization. The aim is to construct such code sequences which can provably achieve the symmetric capacity $C(W)$ for a given B-DMC W, by using the high capacity channels to transmit the *information* bits and using the low capacity channels to transmit the *frozen* bits, i.e. bits with pre-defined values, knowledge of which is also available at the decoder (receiver) and is used during the decoding process as *a priori* knowledge. Depending on the coderate, the channels with the highest capacities after polarization are used to encode *information* bits while those with lowest capacities are used to encode *frozen* bits. The idea of channel polarization is to segregate/partition all the bit channels based on the total available capacity, which is determined by the underlying channel used.

For a binary input RV $X \in \{0, 1\}$ and corresponding output RV Y w.r.t. a given channel, the corresponding conditional PDFs are used to denote the amount of information content and the hence the channel capacity. For B-DMC channels, the mutual information is denoted as [6]:

$$I(W) = \sum_{y \in Y} \sum_{x \in X} W_{X,Y}(x,y) \log_2 \left(\frac{W_{X,Y}(x,y)}{W_X(x)W_Y(y)} \right) \quad (1)$$

and channel capacity is the maximum mutual information of the channel [6]:

$$C(W) = max\left(I(W) \right) \quad (2)$$

If the aforementioned B-DMC is a symmetric channel then (1) can be modified to denote the symmetric $\left(W(x = 0) = W(x = 1) = 0.5 \right)$ channel capacity as follows [1]:

$$I(W) = \tfrac{1}{2} \sum_{y \in Y} \sum_{x \in X} W_{Y|X}(y|x) \log_2 \left(\frac{W_{Y|X}(y|x)}{\frac{1}{2}W_{Y|X}(y|x=0)+\frac{1}{2}W_{Y|X}(y|x=1)} \right) \quad (3)$$

The Bhattacharyya parameter is used to measure the similarity i.e. correlation between different distributions. For any B-DMC it is denoted as [6]:

$$Z(W) \triangleq \sum_{y \in Y} \sqrt{W_{Y|X}(y|x = 0)W_{Y|X}(y|x = 1)} \quad (4)$$

As the PDFs of the channel RVs are always in the range $[0, 1]$, it is easily deduced from (3) and (4) that both $I(W)$ and $Z(W)$ will also be within $[0, 1]$. It is seen from (3) and (4), that I and Z have an inversely proportional relation (terms in the denominator of (3) are the terms in numerator of (4)). Z can thus be used as a parameter to determine the phenomenon of channel polarization as it is an indicator of the channel capacity, i.e. high value of $Z(W)$ indicates low capacity and vice versa.

Suppose channel polarization transforms two copies of channel W into two channels, one with lower (W^-) and the other with higher (W^+) capacity, then the condition $I(W^-) + I(W^+) = 2I(W)$ always holds true and the total capacity is preserved under Shannon's Theorem. This process is shown in Fig. 1.
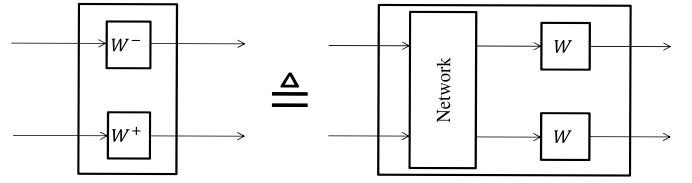

Fig. 1. Black-box depiction of the atomic circuit.

For a 2-bit channel transformation $(W, W) \to (W^-, W^+)$ as shown in Fig. 1, a circuit for channel polarization, is designed such that [1]:

$$I(W^-) + I(W^+) = 2I(W) \quad (5)$$
$$I(W^-) \le I(W) \le I(W^+) \quad (6)$$

Similarly w.r.t. the Z parameters for Fig. 1, following properties hold [1]:

$$Z(W^+) = Z(W)^2 \quad (7)$$
$$Z(W^-) \le 2Z(W) - Z(W)^2 \quad (8)$$
$$Z(W^-) \ge Z(W) \ge Z(W^+) \quad (9)$$

From (7) and (8), it implies that,

$$Z(W^-) + Z(W^+) \le 2Z(W) \quad (10)$$

This idea of channel polarization can also be generalized to N (with $N = 2^n$, where n is a positive integer) independent copies of W channels, in order to synthesize another set of N channels $W_N^{(i)} : 1 \le i \le N$ such that, as N becomes large, the fraction of indices i for which $I(W_N^{(i)}) \approx 1$ approaches $I(W)$ and the fraction of indices i for which $I(W_N^{(i)}) \approx 0$ approaches $1 - I(W)$ while preserving the conditions [1]:

$$\sum_{i=1}^{N} I(W_N^{(i)}) = NI(W) \quad (11)$$

$$\sum_{i=1}^{N} Z(W_N^{(i)}) \le NZ(W) \quad (12)$$

Within the scope of this paper, (7) and (8) are used for $Z(W) \in (0, 1)$. This is because $Z(W) = 0$ means it is a noiseless channel and $Z(W) = 1$ means it is a completely noisy channel, both cases are not applicable for real channels and such channels cannot be polarized to channels with lower/higher capacities, because even after polarization all the channels would have $Z(W) = 0$ or $Z(W) = 1$ effectively resulting in no polarization.

2) Polar Encoding: One way to obtain the aforementioned channel polarization is shown in Fig. 2, which is encapsulated by the "Network" box in Fig. 1. Thus, w.r.t. Fig. 2 for input vector $\boldsymbol{u} = [u_1, u_2]$ for $u_1, u_2 \in \{0, 1\}$, the output vector $\boldsymbol{c} = [c_1, c_2]$ for $c_1, c_2 \in \{0, 1\}$ is generated as follows:

$$c_1 = u_1 \oplus u_2 \quad (13)$$
$$c_2 = u_2 \quad (14)$$

Using Fig. 2, Fig. 1 can be modified to Fig. 3. Fig. 2 along with (13) and (14) are used to provide the matrix representation of the channel polarizing circuit given by (15), which is the generator matrix for encoding a codeword of size $N = 2$ and is the transpose representation of the matrix provided in [1].

$$F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad (15)$$
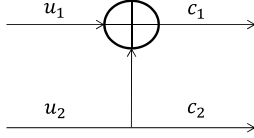

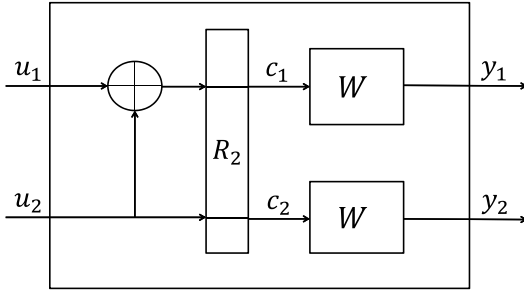Fig. 2. $2 \rightarrow 2$ bit atomic circuit for channel polarization [1].

Fig. 3. 2-bit channel transformation [1].

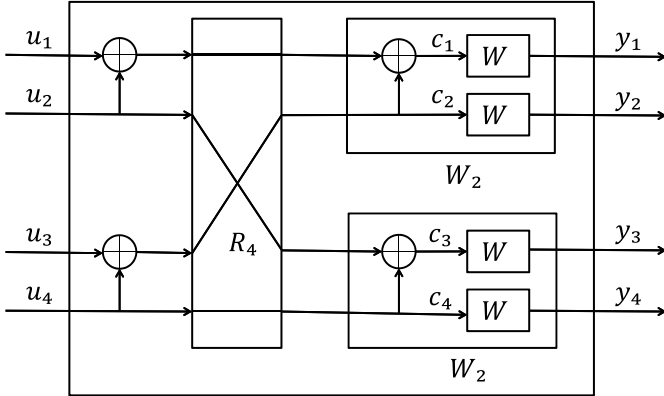A similar channel transformation for $N = 4$ is shown in Fig. 4.

Fig. 4. 4-bit channel transformation [1].

The connections to the circuits in every consecutive stage need to be permuted such that channel transformation using (7) and (8) is possible, i.e. two identical channels with $Z(W)$ are polarized to channels with $Z(W^-)$ and $Z(W^+)$ respectively. The same structure can be generalized to a N-bit circuit such that $N = 2^n$. Consequently, the generator matrix can be represented by [1],

$$G_N = B_N \cdot F^{\otimes n} \qquad (16)$$

where, $B_N \triangleq R_N(I_2 \otimes R_{N/2})(I_4 \otimes R_{N/4}) \cdots (I_{N/2} \otimes R_2)$ is the permutation matrix which creates the desired connections and $\otimes$ is a Kronecker product. $F^{\otimes n}$ is the Kronecker product of the matrix provided by (15), with itself of the order of $n = \log_2 N$. Hence, for an input $\boldsymbol{u}_N$ (consisting of K *information* and $N - K$ *frozen* bits), the codeword $\boldsymbol{c}_N$ is generated by (17) using (16).

$$\boldsymbol{c}_N^\top = G_N \cdot \boldsymbol{u}_N^\top = \left(B_N \cdot F^{\otimes n}\right) \cdot \boldsymbol{u}_N^\top \qquad (17)$$

where, $\boldsymbol{u}_N = \pi_u(\boldsymbol{u}_K \ \boldsymbol{u}_{N-K})$, with $\boldsymbol{u}_K$ being the *information* bit vector, $\boldsymbol{u}_{N-K}$ being the *frozen* bit vector and π_u is a function which maps the bits w.r.t. the bit channel capacities after channel polarization.

3) Polar Decoding: As previously mentioned, to integrate Polar Codes in a BICM-ID design, a SISO decoder is required for exchanging extrinsic information amongst the constituent modules in an ID chain. Amongst, the proposed techniques for polar decoding, Belief Propagation (BP) decoding mentioned by Arikan in [7] is a valid candidate.
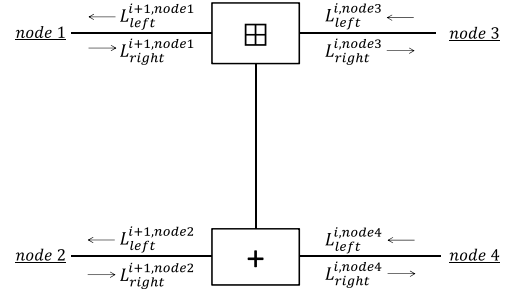

Fig. 5. Mathematical representation of Fig. 2 for BP decoding.

The circuit structure of the decoder is inherently the same as the encoder. Fig. 5 shows the LLR value calculation at each node of an atomic circuit. Comparing Fig. 5 to Fig. 2, the XOR connection results in a boxplus operation, while the direct connection is a simple addition operation of the LLR values. L_{left} is used to denote the *left*-going LLR values, while L_{right} is used to denote the *right*-going LLR values.

$$L_{\text{left}}^{i+1,node1} = L_{\text{left}}^{i,node3} \boxplus \left(L_{\text{left}}^{i,node4} + L_{\text{right}}^{i+1,node2}\right) \quad (18)$$

$$L_{\text{left}}^{i+1,node2} = L_{\text{left}}^{i,node4} + \left(L_{\text{left}}^{i,node3} \boxplus L_{\text{right}}^{i+1,node1}\right) \quad (19)$$

$$L_{\text{right}}^{i,node3} = L_{\text{right}}^{i+1,node1} \boxplus \left(L_{\text{left}}^{i,node4} + L_{\text{right}}^{i+1,node2}\right) \quad (20)$$

$$L_{\text{right}}^{i,node4} = L_{\text{right}}^{i+1,node2} + \left(L_{\text{left}}^{i,node3} \boxplus L_{\text{right}}^{i+1,node1}\right) \quad (21)$$

For any atomic circuit comprising of arbitrary stages i and $i + 1$ $\left(i \in [0, n - 1]\right)$, the LLR values are calculated using (18), (19), (20) and (21) which are the *log*-domain notations of Likelihood Ratio calculations given in [7].

B. BICM-ID

The idea of combining modulation and coding as a single and co-dependent process to improve error correcting performance of the channel coding techniques was proposed in [2]. This approach of coded-modulation was modified, by interleaving

the coded bits to increase code diversity [3]. For the BICM-ID design w.r.t. this paper, multiple codewords are serially/parallely concatenated with the bits of each codeword interleaved with those of the other concatenated codewords, to generate a pseudo-random interleaving. Bits of the same codeword, if interleaved such that they are far apart enough in time to exceed the coherence time, then it would result in different effects of the channel on the bits of the same codeword hence increasing diversity. This provides error protection based not on constraint length (maximum length of codeword, N), but with increased correlation amongst multiple codewords and decreased correlation amongst bits of the same codeword. Ideally maximum diversity can be achieved if the codewords are interleaved such that every bit of the same codeword are transmitted at separate sets of coherence time. Increasing channel diversity is a desired property for a majority of the real-time channels.

The constraint length of the codeword, determines if bit-interleaving would be useful, as the length should be large enough such that transmission time of the entire codeword is longer than the coherence time. This technique of BICM is particularly helpful in case of fading channels, or channel models in which there exist higher degrees of uncertainty/disturbance.
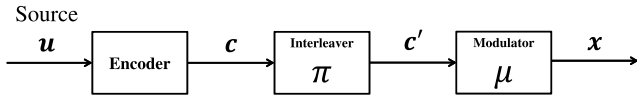

Fig. 6. Block diagram of a BICM design at the transmitter end.

A BICM transmitter is shown in Fig. 6. For the source producing a vector $\boldsymbol{u}$ of K *information* bits and $N-K$ *frozen* bits, it is encoded to a codeword $[c_1, c_2, ..., c_N]$ of length N for a given coderate of K/N of the encoder. A number (say D) of such codewords are concatenated to produce a block $\boldsymbol{c} = [c_{11}, c_{12}, ..., c_{1N}, c_{21}, c_{22}, ..., c_{2N}, ..., c_{D1}, c_{D2}, ..., c_{DN}]$ of length $D \times N$. $\boldsymbol{c}$ is then bit interleaved by a bit interleaver π to produce $\boldsymbol{c}'$. π is a one-to-one correspondence $\pi : i \to i'$, which maps bit at position i to bit at position i', i.e. $\pi(i) = i'$, thus resulting in a time re-ordering of the coded sequence $\boldsymbol{c}$ to produce the sequence $\boldsymbol{c}'$, i.e. $c'_i = c_{\pi(i)}$. The bits of $\boldsymbol{c}'$ are then converted to complex channel symbols $\boldsymbol{x}$, depending on the modulation scheme used by the modulator. For an 2^m-ary constellation map χ obtained by the *m-to-one* mapping μ ($\mu : (0,1)^m \to \chi$) of the encoded and interleaved bits, (22) holds.

$$x_t = \mu(\boldsymbol{c}'_t) \tag{22}$$

where, $\boldsymbol{c}'_t = [c'_{t,1}, c'_{t,2}, \cdots, c'_{t,m}]$ is a set of m bits of the interleaved codewords at an instance t for $c'_{t,i} = c'_{(t-1)m+i}$. Number of instances t would depend on the constellation map used for modulation/demodulation m, length of a codeword N and the number of codewords concatenated for interleaving D i.e. $t \in [1, \frac{D \times N}{m}]$ and $\frac{D \times N}{m} \in \mathbb{N}$.

$x_t \in \chi$ is the modulation symbol obtained by modulating $\boldsymbol{c}'_t$ using (22). These symbols x_t are then transmitted across the channel and the corresponding received output y_t is:

$$y_t = a_t x_t + n_t \tag{23}$$

where, for an AWGN channel, n_t is the additive noise term with Gaussian distribution (with spectral density $N_0/2$) and the attenuation/fade coefficient $a_t = 1 \ \forall t$.
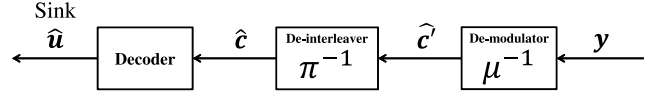

Fig. 7. Block diagram of a BICM design at the receiver end.

A BICM receiver is shown in Fig. 7. The decoder module in the receiver is based on Maximum-Likelihood (ML) decoding as provided in [5]. The ML decoder uses the Free Euclidean Distance (FED) between the transmitted and received symbols to estimate the received symbols and consequently the received bits.

Bit interleaving introduces an additional random modulation causing a reduction in the minimum FED of the received symbols, which might degrade performance over an AWGN channel [3]. To overcome this limitation, for a 2^m-ary modulation scheme, if $m-1$ bits are known by ideal feedback then the corresponding modulation is simplified to a binary modulation of the unknown bit position thus significantly increasing the FED of binary modulation w.r.t. the specific bit position. This is the basis of ID technique, i.e. more reliable bits are used to improve estimation of less reliable bits iteratively as shown in Fig. 8.
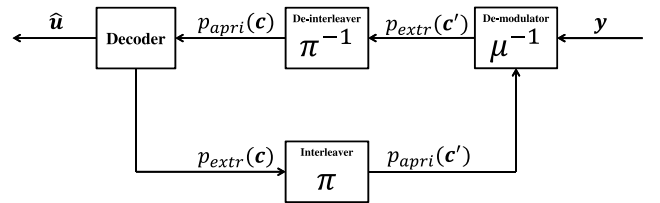

Fig. 8. Block Diagram of ID chain for a BICM-ID design.

For ID, on receiving the channel symbols, the demodulator calculates maximum a posteriori bit metrics corresponding to $\boldsymbol{c}'$. This is the extrinsic information generated by the demodulator. The bit metrics are then deinterleaved and provided to the decoder. The ML-based decoder estimates the input to generate the decoded output. Additionally, it also produces extrinsic information, bit metrics corresponding to which are interleaved to produce a priori information for the demodulator, which are used for demodulating the channel symbols again in the next iteration of ID. This entire process is iterated as many times as is required to converge to a solution.

III. BIPCM-ID

The novel BIPCM-ID system is developed by using Polar Codes as the underlying error correcting codes in a BICM-ID design which is shown in Fig. 9. Clearly, the conventional
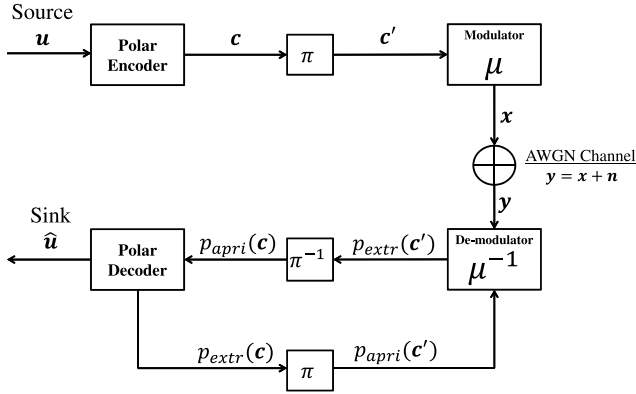

Fig. 9. Block Diagram of a BIPCM-ID design.

encoder and decoder modules are replaced by the polar encoder and decoder modules, respectively. Polar encoding is performed exactly as described in Section II.

The decoder however is the trickier module to be integrated within the design. As previously discussed, a SISO decoder is used to generate soft LLR values for ID, which in this case would be the BP decoder as explained in Section II. To be able to use the BP decoder for ID, LLR values corresponding to the extrinsic a posteriori probabilities $p_{extr}(c)$ need to be calculated which would be interleaved and forwarded to the demodulator for a consecutive iteration. Thus, extrinsic information should also be an output generated by the decoder for every iteration of the ID. The desired decoder structure is shown in Fig. 10.
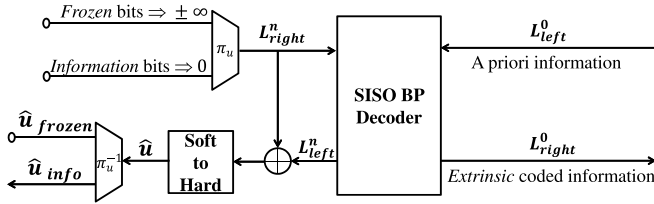

Fig. 10. Polar decoder in a BIPCM-ID system.

A. Extrinsic Information

At each node of the BP decoder (see Fig. 5), there are 2 LLR values available. One is a *left*-going LLR value while the other is a *right*-going LLR value. Owing to the circuit structure for generating Polar Codes, it is evident that there are $n = log_2 N$ stages within the circuit with every stage having N nodes each. The BP decoder however, is a black box module for the other components in the ID chain at the receiver which means that the interleaver and deinterleaver have access only to the *left* and *right*-going LLR values at the first and last stages w.r.t. the circuit structure of the decoder. As evident from Fig. 9,

the input to the decoder is a set of N *left*-going LLR values. These input LLR values would be represented as $\boldsymbol{L}_{\text{left}}^0$ i.e. LLR values at the initial stage 0 of the decoder. Thus, the *left*-going LLR values at the decoder output can be represented as $\boldsymbol{L}_{\text{left}}^n$ i.e. LLR values at the last stage n of the decoder. The same convention is used to represent the *right*-going LLR values. i.e. $\boldsymbol{L}_{\text{right}}^n$ are the *right*-going LLR values at the output of the decoder, while $\boldsymbol{L}_{\text{right}}^0$ are the *right*-going LLR values at the input of the decoder.

$\boldsymbol{L}_{\text{left}}^0$ consists of the LLR values corresponding to the codeword to be decoded, provided as input to the decoder. Thus,

$$L_{\text{left}}^{0,i} = L_{\hat{c}_i} = \log_e \frac{p_{apri}(c_i = 0)}{p_{apri}(c_i = 1)} \quad (24)$$

and,

$$\boldsymbol{L}_{\text{left}}^0 = \left[L_{\text{left}}^{0,1}, L_{\text{left}}^{0,2}, L_{\text{left}}^{0,3}, \cdots L_{\text{left}}^{0,N} \right] \quad (25)$$

$\boldsymbol{L}_{\text{right}}^n$ consists of the a priori knowledge of *frozen* bits and they are set such that the LLR value calculations (as provided in Sec. II-A3) in the decoder should not affect the estimate of the *frozen* bits as well as help in calculating LLR values corresponding to *information* bits. Thus,

$$L_{\text{right}}^{n,i} = \log_e \frac{p(u_i = 0)}{p(u_i = 1)} = \begin{cases} +\infty, & \text{if } u_i \text{ is a } frozen \text{ bit} = 0 \\ -\infty, & \text{if } u_i \text{ is a } frozen \text{ bit} = 1 \\ 0, & \text{if } u_i \text{ is not a } frozen \text{ bit} \end{cases} \quad (26)$$

and

$$\boldsymbol{L}_{\text{right}}^n = \left[L_{\text{right}}^{n,1}, L_{\text{right}}^{n,2}, L_{\text{right}}^{n,3}, \cdots L_{\text{right}}^{n,N} \right] \quad (27)$$

The LLR values at the input to the decoder belong to the same domain as the output of the encoder which is evident from Fig. 9. The LLR values $\boldsymbol{L}_{\text{left}}^0$ and $\boldsymbol{L}_{\text{right}}^0$ at the input to the decoder should contain maximum amount of information about the codeword. $\boldsymbol{L}_{\text{right}}^0$ correspond to the set of *right*-going LLR values which are calculated at the last instance, i.e. *right*-going LLR values are calculated from stage n to stage 0. Thus, intuitively $\boldsymbol{L}_{\text{right}}^0$ should contain maximum amount of extrinsic information from the decoder to be used for ID, as its values would reflect decoding effects from all the stages of the decoder. The simulation results (discussed in the following section) have been generated by using $\boldsymbol{L}_{\text{right}}^0$ and the Bit Error Rate (BER) characteristics clearly show that $\boldsymbol{L}_{\text{right}}^0$ is the proper choice for extrinsic information.

The LLR values at the decoder output as shown in Fig. 10, are used to estimate the input. From, Fig. 9, it is evident that the leftmost LLR values $\boldsymbol{L}_{\text{left}}^n$ and $\boldsymbol{L}_{\text{left}}^n$ correspond to the same domain as input to the encoder, which are the bits that need to be estimated. Hence, w.r.t. Fig. 10, the output LLR values used to estimate the input vector is denoted by (28).

$$\boldsymbol{L}_{\text{out}} = \boldsymbol{L}_{\text{right}}^n + \boldsymbol{L}_{\text{left}}^n \quad (28)$$

where, $\boldsymbol{L}_{\text{right}}^n \in \{0, +\infty, -\infty\}$. These LLR values are soft decision values which can be transformed to hard decision values by the following method:

$$\hat{u}_i = \begin{cases} 0, & \text{if } L_{\text{out},i} \geq 0 \\ 1, & \text{if } L_{\text{out},i} < 0 \end{cases} \quad (29)$$

B. Channel Polarization

Arikan proposes to use the value of Z to polarize channels. However his method of polarizing channels is specific only to BEC. The aim in this paper is to develop a BIPCM-ID system over an AWGN channel. So, if W is a Binary Input AWGN (BI-AWGN) channel with input $X \in \{+1, -1\}$ ($x = +1 \implies$ bit 0 and $x = -1 \implies$ bit 1), (4) can be modified as [6],

$$Z(W) = \int_{\forall y \in Y} \sqrt{W_{Y|X}(y|x=-1) \cdot W_{Y|X}(y|x=+1)} dy \tag{30}$$

where,

$$W_{Y|X}(y|x=+1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-1)^2}{2\sigma^2}} \ ,$$

$$W_{Y|X}(y|x=-1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y+1)^2}{2\sigma^2}} \ .$$

Solving (30) by substituting the corresponding values of $W_{Y|X}$ gives [9]

$$Z(W) = e^{-\frac{1}{2\sigma^2}} = e^{-S/N} \tag{31}$$

for

$$\frac{S}{N} = \frac{E_s}{N_0} = \frac{E_b}{N_0} \cdot R_{mod} \cdot R_c$$

where E_s/N_0 is the energy per symbol to noise power spectral density ratio, E_b/N_0 is the energy per bit to noise power spectral density ratio, R_{mod} is the number of bits in one symbol and R_c is the coderate of the channel. Evidently, from (31), it can easily be concluded that $Z(W)$ used for channel polarization of an AWGN channel is a function of the channel's Signal to Noise Ratio. Thus, Channel State Information (CSI) of an AWGN channel can be exploited to polarize channels and using (31), $Z(W)$ can be updated when value of E_b/N_0 changes. However CSI of the AWGN channel may not always be available at the transmitter. In such a situation the value of $Z(W)$ (to polarize channels) is once determined and kept unchanged for a certain design setting throughout all channel conditions (i.e. all values of E_b/N_0 over a specific N).

BER simulations have shown that using CSI to polarize channels do not provide optimal BER performance and better BER can be achieved by fixing $Z(W)$ to polarize channels in all channel conditions. Consequently, within the scope of this paper, CSI is not used for channel polarization. As discussed already in Sec. II-A2, values of E_b/N_0 are considered such that $Z(W) \approx 0$ or $Z(W) \approx 1$ are avoided. For this paper, E_b/N_0 in the range $[-5\,\mathrm{dB}, 10\,\mathrm{dB}]$ have been considered to polarize channels using (31) and the corresponding BER curves have been analysed.

Fig. 11 shows the range of values of $Z(W)$ which provides the optimal BER performance corresponding to encoding and decoding of Polar Codes, over multiple values of N and different modulation schemes with Gray mappings for $R_c = 1/2$ over AWGN channel. The optimal BER performance is determined by the curve with the lowest BER in the waterfall region of the BER curve, i.e. the region where the BER curve tends to vanishingly low values (of the order of 10^{-4} or lower). For a given N in Fig. 11, the leftmost bar depicts the optimal
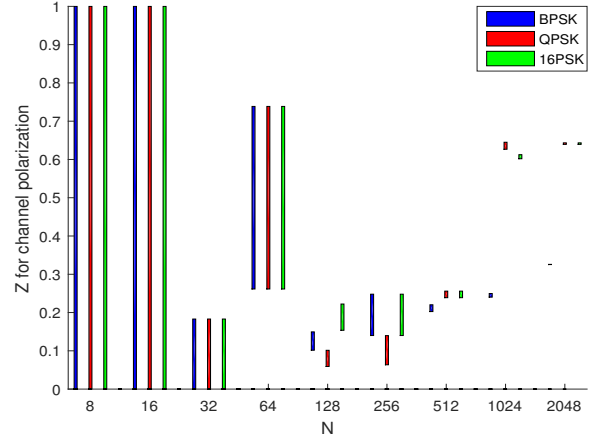

Fig. 11. Z values providing optimal BER performance for given N and modulation scheme.

values of $Z(W)$ for Binary-Phase Shift-Keying (BPSK), the center bar for Quadrature-Phase Shift-Keying (QPSK) and the rightmost bar for 16-Phase Shift-Keying(16PSK) modulation schemes.

From Fig. 11 it can be concluded that for $N = 128$,

1) While using the BPSK modulation scheme, polarizing channels with $Z(W) \approx 0.1066$ gives the optimal BER performance. From Fig. 11 it is evident that any value in the range $\forall Z(W) \in [0.1015, 0.1495]$ (denoted by Z^{128}_{BPSK}) would provide the same optimal BER performance.

2) While using the 16PSK modulation scheme, polarizing channels with $Z(W) \approx 0.1689$ gives the optimal BER performance. From Fig. 11 it is evident that any value in the range $\forall Z(W) \in [0.1536, 0.2223]$ (denoted by Z^{128}_{16PSK}) would provide the same optimal BER performance.

The aforementioned values have been determined for the condition when Polar Codes are the only error correction method. However, with the parameters used, the same comparative BER performance is applicable for a BIPCM design as bit-interleaving and coded-modulation framework is unaffected/would not affect i.e. independent of the process of channel polarization. These values of $Z(W)$ are used to design a BIPCM system which in turn would be the benchmark for comparing performance of the novel BIPCM-ID system, simulation results of which are provided in the following section.

IV. SIMULATION RESULTS

For the novel BIPCM-ID system, Error Free Feedback (EFF) is obtained by ideal feedback which is available at the receiver, i.e. a priori information to the demodulator (bit metrics corresponding to $p_{apri}(c)$ in Sec. II-B) corresponding to the bits generated after interleaving (as in Fig. 6) and provided by the transmitter. The EFF results for the BIPCM-ID has been obtained from simulations.

For EFF, the 2^m-ary constellations are converted to binary signal labeling (equivalent to BPSK modulation) amongst 2^{m-1} pairs. Although using Gray mapping provides smallest distance between 1-bit neighbors (w.r.t. the constellation map), the inter-symbol FED for a pair of binary labeling for ideal feedback (EFF) remains unchanged from the original constellation map. Thus, for a non-ID scheme, Gray mapping would be the best choice for symbol mapping. However, for ID schemes, as iterations would not be able to change the binary signal labeling, using Gray mapping would not help in improving error performance over successive iterations within ID. If a constellation map is selected such that the Harmonic Mean of the minimum squared FED is effectively increased, then there is scope for improvement of error correcting performance over successive iterations within ID. A SSP map provides such desired characteristics and is hence used for mapping symbols in the ID scheme.

The BIPCM-ID system has been analyzed with the following values of the system parameters:

1) $N = 128$.
2) Number of codewords concatenated for interleaving $D = 1600$. Thus, the blocks $\boldsymbol{c}$, $\boldsymbol{c'}$, $\hat{\boldsymbol{c}}$ and $\hat{\boldsymbol{c}}'$ are of size $D \times N = 1600 \times 128 = 204800$ bits.
3) Coderate = $1/2$.
4) Gray mapping used for non-ID scheme (BIPCM) and SSP mapping used for ID scheme (BIPCM-ID).
5) The LLR value calculation within the BP decoder is done iteratively with one stage at a time and both the *left* and *right* going LLR values are calculated over every iteration. 60 iterations are used for calculating the LLR values w.r.t. one codeword, for converging to the solution.

As Polar Codes have been designed with the constraint $N = 2^n$, only 2^m-ary schemes would be used for mapping the symbols. The BER performance of a BIPCM-ID system for 16-PSK modulation is shown in Fig. 12.
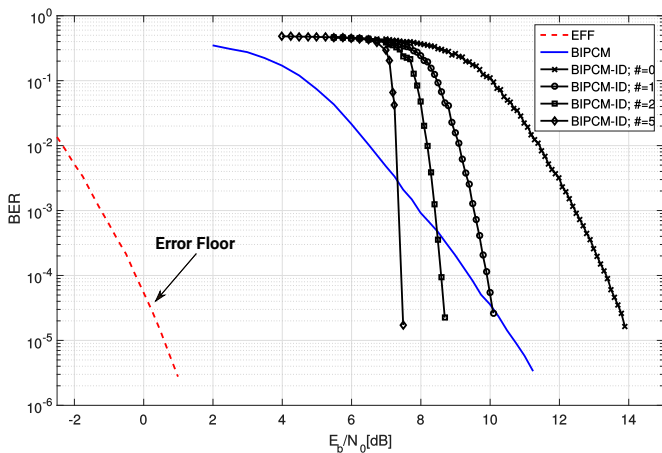

Fig. 12. BER performance of BIPCM-ID, with 16-PSK modulation scheme.

Owing to the results shown in Fig. 11 and by (31), using Z_{16PSK}^{128} for channel polarization over 16PSK modulation

scheme provides the optimal BER performance for non-ID usage of Polar Codes, which in this case is the BIPCM system with 16PSK Gray constellation map marked by the blue curve in Fig. 12. It is the optimal performance achievable by the BIPCM system and is thus the benchmark over which BIPCM-ID system performance is assessed. The red curve represents the EFF performance of the BIPCM-ID system indicating its performance limit and providing the corresponding error floor. For a BIPCM-ID system, using 16PSK SSP mapping and $Z(W) = 0.5$, provides the optimal BER performance and the performance improvement over increasing number of iterations (indicated by $\# = 0$, 1, 2 and 5) is shown in Fig. 12. As expected from any ID scheme, BER performance of the BIPCM-ID system improves with increasing number of iterations and the amount of improvement over consecutive iterations reduces for higher number of iterations. BIPCM-ID with at least 5 iterations clearly outperforms BIPCM at $E_b/N_0 \approx 7.3$ dB beyond the BER range of the order of 10^{-3} and it achieves vanishingly small BER (of the order of 10^{-5}) at $E_b/N_0 \approx 7.5$ dB and beyond. This proves that a BIPCM-ID system can be designed which can outperform the corresponding BIPCM system.

The least complex modulation scheme is BPSK. However, BPSK is inherently a Gray mapping scheme with fixed FED of the constellation map with no alternate mappings available. This makes using BPSK suitable for a BIPCM system but not for BIPCM-ID. The analysis provided in [10] shows that all known constellation maps for any modulation scheme would inherently have an error floor in a BICM-ID design, beyond which the error performance cannot be improved inspite of using a very high number of iterations for ID. The idea proposed in [10] is to introduce *differential* encoding/decoding to remove this error floor. Using a Differential Binary Phase Shift-Keying (DBPSK) modulation scheme would thus not only remove error floor of the system, but also provide a modulation scheme for BIPCM-ID system which can be compared to the BIPCM system over BPSK, with no added code redundancy but little additional complexity. Fig. 13 shows the corresponding comparative BER performance.

Referring to Fig. 11 and (31) Z_{BPSK}^{128} is used for channel polarization to achieve optimal performance of the BIPCM system. The blue BER curve in Fig. 13, shows the optimal BER performance of BIPCM with BPSK modulation scheme and is the benchmark to compare performance of the corresponding BIPCM-ID system. The BIPCM-ID system designed with DBPSK modulation scheme has no error floor and $Z(W) = 0.5$, provides the optimal BER performance. Behaviour of performance improvement for increasing number of iterations (indicated by $\# = 0$, 1, 2 and 5) is as expected (similar behaviour as with using 16PSK modulation). BIPCM-ID with at least 5 iterations clearly outperforms BIPCM at $E_b/N_0 \approx 2$ dB beyond the BER range of the order of 10^{-2}. Thus, DBPSK modulation can be used to design a BIPCM-ID system without an error floor, which can outperform the corresponding BIPCM system, hence resulting in a high performance system with vanishingly small BER (of the order of 10^{-6}) at low values
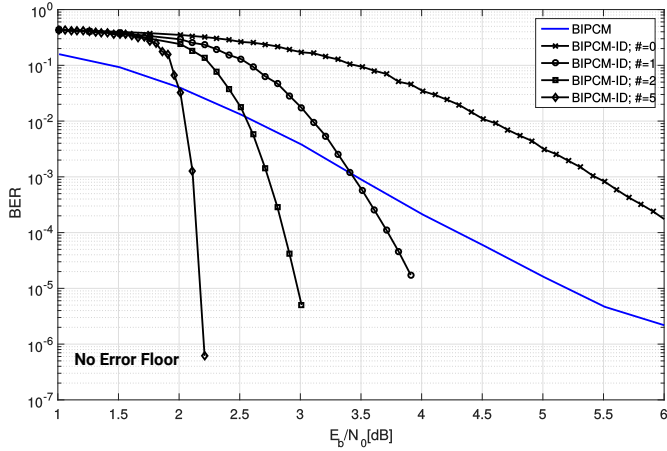
Fig. 13. BER performance of BIPCM-ID, with BPSK/DBPSK modulation schemes.

of $E_b/N_0 \approx 2.2$ dB and beyond.

The aforementioned results thus prove that it is possible to design a BIPCM-ID system over different 2^m-ary modulation schemes which can outperform a corresponding BIPCM system.

V. FUTURE WORK

The BER performance of the BIPCM-ID system developed provide promising results. Nevertheless, a number of challenges have been encountered which have risen some unanswered questions, that if solved would not only improve the existing system but would also help to achieve higher throughput. Following points reflect the main points which are prospective areas for future research:

1) BIPCM-ID system analysis for larger codewords (higher values of N).
2) EXIT Chart analysis of the BIPCM-ID system to determine the parameter settings for optimal BER performance.
3) Performance analysis of using a less complex Polar Decoder (like SCAN algorithm) in the BIPCM-ID system or using a less complex way of calculating the LLR. The Belief Propagation (BP) Polar Decoder in the BIPCM-ID system is the module with highest computational complexity and it uses $tanh$ and exp functions to calculate the Logarithmic Likelihood Ratio (LLR) values using boxplus operation. Using an easier calculation method like $min - sum$ would drastically reduce the complexity of BP decoder, hence effectively reducing the complexity of BIPCM-ID system.
4) Developing a way to polarize channels, to generate optimal choice of Polar Codes if no CSI is available.
5) Analyzing the relationship between $Z(W)$ for channel polarization and the coderate.
6) Analyzing the performance of a BIPCM-ID system for $N \neq 2^n$. With $N \neq 2^n$, non-2^mary constellations (e.g. 8PSK) can be used for modulation/demodulation.

7) Analyzing the BER performance of BIPCM-ID system over other channel models (e.g. Fading channels).
8) Determining the number of iterations required by BIPCM-ID for convergence of the BER performance.

VI. CONCLUSION

A novel BIPCM-ID system has been developed using a polar encoder (at the transmitter) and a polar decoder (at the receiver) within a BICM-ID design. This system has been analyzed over AWGN channels with $N = 128$ for coderate$= 1/2$ over BPSK and 16PSK modulation schemes. Implementing the system with these parameters, it has been observed that under proper configurations (parameter values),

1) with 5 iterations of ID, BER performance can be improved from its contemporary BIPCM system by at least 3.8 dB over BPSK modulation for very low BER.
2) with 5 iterations of ID, BER performance can be improved from its contemporary BIPCM system by at least 3 dB over 16PSK modulation for very low BER.

High code diversity is very important to be able to well utilize the potential of a BIPCM-ID system. With a higher number of codewords concatenated for bit interleaving, higher code diversity and lower correlation amongst transmitted bits of the same codeword is achieved.

The performance of BIPCM-ID can further be improved by increasing the number of ID iterations. Using a differential modulation scheme removes the error floor altogether.

Polar Codes as a stand-alone error correction technique are provably capacity achieving codes for B-DMCs especially BEC. They are not specifically designed for high performance error correction over continuous channels. However, by developing a BIPCM-ID system it has been proved that with the help of some additional error correcting modules, Polar Coding can be a high performance error correction method over continuous channels like AWGN as well.

REFERENCES

[1] E. Arikan, "Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051-3073, Jul. 2009.
[2] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 55-67, Jan. 1982.
[3] E. Zehavi, "8-PSK Trellis Codes for a Rayleigh Channel," *IEEE Transactions on Communications*, vol. 40, no. 5, pp. 873-884, May. 1992.
[4] G. Caire, G. Taricco and E. Biglieri, "Bit-Interleaved Coded Modulation," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 927-946, May. 1998.
[5] X. Li and J. Ritcey, "Bit-Interleaved Coded Modulation with Iterative Decoding," *IEEE Communications Letters*, vol. 1, no. 6, pp. 169-171, May. 1997.
[6] J.G. Proakis and M.Salehi, *Digital Communications*. McGraw Hill, 1221 Avenue of the Americas, New York, NY 10020, ed. 5, 2008.
[7] E. Arikan, "Polar codes: A pipelined implementation," *Proc. 4th ISBC*, pp. 11-14, Jul. 2010.
[8] X. Li, A. Chindapol and J. Ritcey, "Bit-Interleaved Coded Modulation with Iterative Decoding and 8PSK Signalling," *IEEE Transactions on Communications*, vol. 50, no. 8, pp. 1250-1257, Aug. 2002.
[9] H. Li and J. Yuan, "A practical construction method for polar codes in AWGN channels," *IEEE 2013 Tencon - Spring*, pp. 223-226, April. 2013.
[10] S. Pfletschinger and F. Sanzi, "Error Floor Removal for Bit-Interleaved Coded Modulation with Iterative Detection," *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 3174-3181, Nov. 2006.