Polar Codes for Magnetic Recording Channels

Aman Bhatia*, Veeresh Taranalli*, Paul H. Siegel*, Shafa Dahandeh†,
Anantha Raman Krishnan†, Patrick Lee†, Dahua Qin†, Moni Sharma†, and Teik Yeo†
*University of California, San Diego, La Jolla, CA, USA
Email: {amanbh, vtaranalli, psiegel}@ucsd.edu
†Western Digital Corporation, CA, USA
Email: {Shafa.Dahandeh, Anantha.Krishnan, Patrick.Lee, Dahua.Qin, Moni.Sharma, Teik.Yeo}@wdc.com

***Abstract*—Polar codes provably achieve the capacity of binary memoryless symmetric (BMS) channels with low complexity encoding and decoding algorithms, and their finite-length performance on these channels, when combined with suitable decoding algorithms (such as list decoding) and code modifications (such as a concatenated CRC code), has been shown in simulation to be competitive with that of LDPC codes. However, magnetic recording channels are generally modeled as binary-input intersymbol interference (ISI) channels, and the design of polar coding schemes for these channels remains an important open problem. Current magnetic hard disk drives use LDPC codes incorporated into a turbo-equalization (TE) architecture that combines a soft-output channel detector with a soft-input, soft-output sum-product algorithm (SPA) decoder. An interleaved coding scheme with a multistage decoding (MSD) architecture with LDPC codes as component codes has been proposed as an alternative to TE for ISI channels. In this work, we investigate the use of polar codes as component codes in the TE and MSD architectures. It is shown that the achievable rate of the MSD scheme converges to the symmetric information rate of the ISI channel when the number of interleaves is large. Simulations results comparing the performance of LDPC codes and polar codes in TE and MSD architectures are presented.**

I. Introduction

Channel polarization was originally proposed in [1] for binary input discrete memoryless channels. This technique allows transforming multiple copies of an arbitrary symmetric binary-input discrete memoryless channel (B-DMC) to create 'bit-channels' that are either noiseless or completely noisy, such that the fraction of bit-channels that are essentially noiseless is close to the capacity of the original B-DMC. Identifying which bit-channels are noiseless after this polarization transformation allows the construction of *polar codes*, a family of codes that provably achieve the capacity of B-DMC with low-complexity encoding and decoding algorithms.

Since construction of polar codes has complexity exponential in blocklength for channels other than the binary erasure channel, a number of approximate construction methods that have linear complexity have been proposed [2], [3], [4]. While polar codes achieve the capacity with a successive cancellation (SC) decoder at infinite blocklengths, the performance of finite-length polar codes can be improved by the use of better decoding algorithms. Successive cancellation list decoder was proposed in [5] for hard-decision decoding and the performance was comparable to that of LDPC codes with belief propagation (BP) decoding. SC-list decoding operates like SC decoding, but whenever a decision on an unfrozen bit is needed, it splits the decoding path into two to try both 0 and 1. When the number of paths grows beyond the prescribed threshold L, the decoder keeps only the L most likely paths. BP decoding over the polar code factor graph was proposed in [6] as a soft-output decoder for polar codes. Soft cancellation (SCAN) decoder, proposed in [7], uses a message passing schedule motivated by the SC decoder and improves the soft-output decoding performance further.

Polar codes are closely related to Reed-Muller (RM) codes, a connection that was identified in [1]. The rows of the generator matrix for both codes are chosen from the rows of the matrix $\left[\begin{smallmatrix}1 & 0\\ 1 & 1\end{smallmatrix}\right]^{\otimes n}$ where $\otimes$ denotes the Kronecker product. For RM codes, the rows with the largest Hamming weight are chosen. However, for polar codes, rows corresponding to the noiseless channels obtained by the recursive channel-polarization transforms are chosen. Thus the choice depends on the channel and requires that the recursive transforms polarize the channel. Magnetic recording channels are modeled as a binary-input channel with finite memory due to intersymbol interference. The polarization phenomenon for stationary, ergodic, finite-order Markov processes through the polarization transform has been shown in [8]. However, it is only conjectured that the polarization result extends to intersymbol interference channels. Also, code construction for channels with memory remains an open problem as the code construction techniques in [2], [3], [4] only apply for memoryless channels. Thus, the construction of capacity-achieving polar codes for magnetic recording channels is a challenging problem.

Current magnetic recording technologies use LDPC codes as error-correction codes (ECC) incorporated into a turbo-equalization (TE) architecture that combines a soft-output channel detector with a soft-input soft-output (SISO) LDPC decoder. An interleaved coding with multistage decoding (MSD) architecture with LDPC codes as component codes has been proposed in [9] as an alternative to TE for ISI channels. In this scheme, M independent codewords are interleaved at the encoder. At the receiver, the M codewords are decoded sequentially, with each decoded codeword being used for decoding subsequent interleaves. It is known that the achievable rate of this MSD scheme converges to the symmetric information rate (SIR) of the ISI channel when $M \to \infty$.

In this work, we investigate the use of polar codes in system architectures based on turbo-equalization and interleaved coding with multistage decoding. First we study the incorporation of polar codes for use with the TE architecture. Since this requires a soft-output polar decoder, neither SC decoding nor

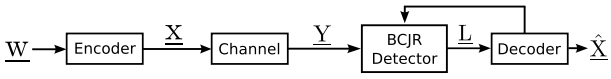

Fig. 1. Encoder and decoder for turbo-equalization.

SC-list decoding can be used. Instead, we need a soft-output variant of the SC decoder, such as the BP decoder, or the SCAN decoder. We compare some methods to construct polar codes for use in this architecture.

We also propose the use of an interleaved coding scheme with multistage decoding with polar codes as component codes. The MSD architecture has several attractive features when polar codes are used as component codes. First, this architecture does not require soft outputs from the component polar decoders, so SC and SC-list decoders may be used. Second, the component polar code for each stage can be designed using techniques developed for memoryless channels. Finally, common hardware designs for encoding and decoding can be shared among the interleaves.

The rest of the paper is organized as follows. The channel model and the TE and MSD architectures are presented in Sections II, III, and IV, respectively. Simulation results are presented in Section V. Conclusions are discussed in Section VI.

II. Channel Model

Magnetic recording channels are modeled as binary-input intersymbol-interference channels with additive white Gaussian noise (AWGN). More specifically, if $\underline{X} = \{x_k\}$ is the input sequence drawn from a binary alphabet $\{\pm 1\}$, the channel output at time n is given by $y_n = \sum_{i=0}^{\nu} h_i x_{n-i} + a_n$. Here $\{h_i\}$ is a known channel impulse response associated with the transfer function $h(D) = \sum_{i=0}^{\nu} h_i D^i$ and a_n is i.i.d. Gaussian noise with zero mean and variance σ^2. Such a channel is referred to as a partial-response (PR) channel. A well known example is the dicode channel which has transfer function $h(D) = 1 - D$. We note here that the results presented in [8] regarding polarization in the presence of memory do not apply to PR channels.

III. Turbo-Equalization Architecture

Current magnetic storage drives use a decoding architecture based upon turbo-equalization [10] that combines a SISO sequence detector and a SISO decoder exchange information iteratively as shown in Figure 1. Here, the user data sequence is denoted by $\underline{W}$, the output of the ECC encoder is denoted by $\underline{X}$, and the channel output sequence is denoted by $\underline{Y}$. To decode the user data, the BCJR algorithm [11] is used to determine the log-a-posteriori probability (APP) ratios $\underline{L}$. An ECC decoder feeds back soft information to the sequence detector until it can decode to the codeword $\hat{\underline{X}}$.

There are two issues in using polar codes as ECC in this architecture. First, we need to use a soft-output decoder to be able to feed back soft information to the sequence detector as priors. While BP and SCAN decoders can provide soft-outputs, a larger coding gain is obtained when polar codes

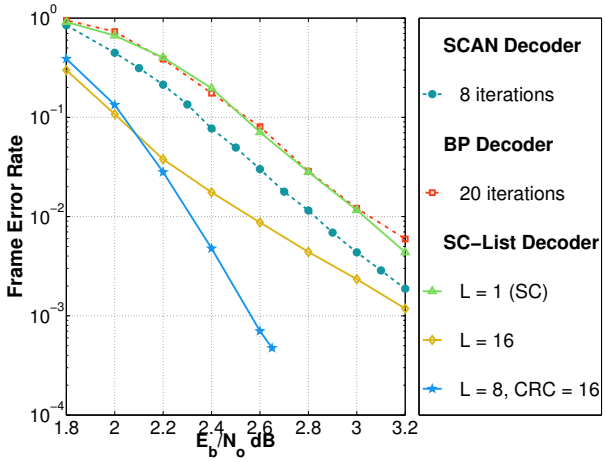

Fig. 2. Performance of a rate-0.7, blocklength-4096 polar code on AWGN channel when decoded using BP, SCAN, SC and SC-list decoders. Concatenation with an outer 16-bit CRC code is used for SC-list with $L = 8$.

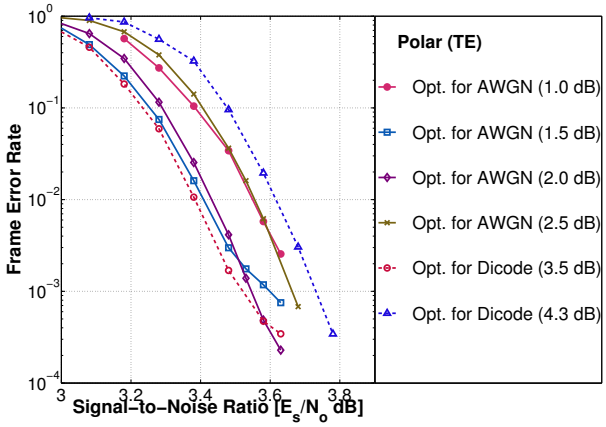

Fig. 3. Performance of polar codes optimized for various AWGN and dicode channels with rate 0.83 (corresponding to design SNR 2.5 dB) and blocklength 16384, decoded using SCAN on dicode channel.

are decoded by the SC-list decoder which only provides hard decisions. This is demonstrated in Figure 2 which shows the performance of the BP, SCAN, SC, and SC-list for a rate 0.7 polar code on a AWGN channel. The second issue is that the choice of unfrozen bit-channels during polar code construction depends on the underlying channel. The equivalent channel seen by the ECC decoder is not memoryless and there is no method known for making the optimal choice. Therefore, we use the following two sub-optimal approaches. In the first approach, we construct polar codes using the method of [3] for AWGN channels with different noise variances. In the second approach, the conditional output density of the equivalent channel seen by the ECC decoder on the first iteration of TE, that is $\mathbb{P}(L_j|C_j)$, is estimated using Monte-Carlo simulation. Polar codes are constructed for this equivalent channel by assuming it to be memoryless (not necessarily true). The simulation results obtained for codes constructed by these two approaches are shown in Figure 3.

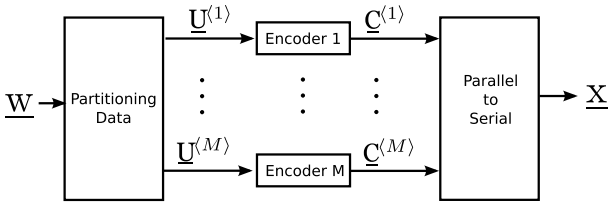

Fig. 4. Encoder for interleaved coding scheme.

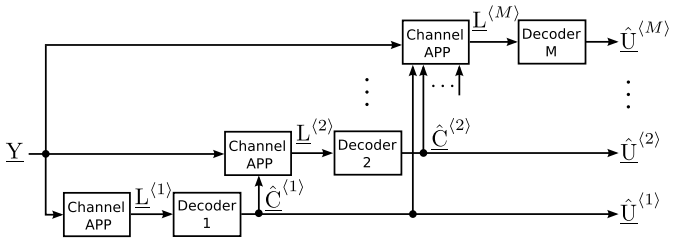

Fig. 5. Multi-stage decoding scheme.

IV. Multistage Decoding Architecture

An *interleaved code* design with a *multi-stage decoding* (MSD) algorithm is considered as an alternative to turbo equalization for binary ISI channels in [9]. In this scheme, M independent codewords of potentially different rates are interleaved at the encoder. At the receiver, each of the M codewords are decoded sequentially, with each decoded codeword being used for decoding subsequent interleaves. Figure 4 shows the block diagram for the encoder. The user data, $\underline{\mathrm{W}}$, is partitioned into M blocks, $\{\underline{\mathrm{U}}^{\langle i\rangle}\}_{i=1}^{M}$ of lengths $K^{\langle 1\rangle}, K^{\langle 2\rangle}, \ldots, K^{\langle M\rangle}$ respectively. Each of these M blocks are encoded using codes $\mathcal{C}_1, \ldots, \mathcal{C}_M$ of rates $R_i = \frac{K^{\langle i\rangle}}{N}$ such that the codewords $\{\underline{\mathrm{C}}^{\langle i\rangle}\}_{i=1}^{M}$ are all of length N bits, denoted as $\left(C_1^{\langle i\rangle}, \ldots, C_N^{\langle i\rangle}\right)$ for each i. These M codewords are then randomly interleaved into a sequence $\underline{\mathrm{X}}$ which is transmitted.

Figure 5 presents the decoder architecture for multi-stage decoding. Given received word $\underline{\mathrm{Y}}$, the decoding proceeds to recover the interleaved codewords in M stages. In stage i, let $\hat{\underline{\mathrm{C}}}^{\langle k\rangle}$ denote the estimate for the k^{th} codeword obtained by the decoder at stage $k = 1, 2, \ldots, i-1$. Then, the channel detector determines the log-APP ratio vector $\underline{\mathrm{L}}^{\langle i\rangle} = \left(L_1^{\langle i\rangle}, L_2^{\langle i\rangle}, \ldots, L_N^{\langle i\rangle}\right)$ according to

$$L_j^{\langle i\rangle} = \log\left(\frac{1-P_j^{\langle i\rangle}}{P_j^{\langle i\rangle}}\right),$$

where,

$$P_j^{\langle i\rangle} = \mathbb{P}\Big(C_j^{\langle i\rangle} = 1 \Big| \underline{\mathrm{Y}}, \underline{\mathrm{C}}^{\langle 1\rangle} = \hat{\underline{\mathrm{C}}}^{\langle 1\rangle}, \ldots \underline{\mathrm{C}}^{\langle i-1\rangle} = \hat{\underline{\mathrm{C}}}^{\langle i-1\rangle}\Big).$$

$\underline{\mathrm{L}}^{\langle i\rangle}$ can be computed efficiently using the BCJR algorithm on the entire received sequence while assuming that the estimates for codeword at previous stages are correct and assuming the codewords belonging to the future stages are distributed independently and uniformly. These log-APP ratios are then used by the i^{th} component decoder to estimate the i^{th} codeword as $\hat{\underline{\mathrm{C}}}^{\langle i\rangle}$.

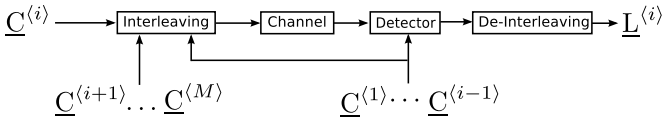

Fig. 6. Equivalent sub-channel for i^{th} stage with $\underline{\mathrm{C}}^{\langle k\rangle}, k = 1, \ldots, i-1$, known to decoder.

The channel, as seen by the code at the i^{th} stage, can be represented by the *equivalent sub-channel* shown in Figure 6. It is shown in [9] that a rate

$$R^{\langle i\rangle} \triangleq I\left(C_j^{\langle i\rangle}; L_j^{\langle i\rangle}\right) = 1 - \mathbb{E}\Big[\log_2\left(1+e^{-L_j^{\langle i\rangle}}\right) \Big| C_j^{\langle i\rangle}\Big] \tag{1}$$

is an achievable information rate for the i^{th} stage, even when the decoder has only the knowledge of the marginal channel law $\mathbb{P}\Big(C_j^{\langle i\rangle}, L_j^{\langle i\rangle}\Big)$. It is shown in [9] that random codes achieve these bounds with decoders that assume that the equivalent sub-channels are memoryless and that use only partial knowledge about the channel in the form of the marginal output densities. The overall rate, $R_{\text{av},M} = \frac{1}{M}\sum_{i=1}^{M} R^{\langle i\rangle}$, is shown to converge to the SIR of the ISI channel as $M \to \infty$. The paper then optimizes component LDPC codes for each sub-channel to achieve thresholds close to the achievable rates.

In this work, we propose the use of polar codes in place of LDPC codes as interleaved codes in the MSD scheme described in [9]. One can simulate the equivalent sub-channel assuming perfect decision feedback from the component decoders to obtain a set of input and output sequences. The densities $\mathbb{P}\Big(C_j^{\langle i\rangle}, L_j^{\langle i\rangle}\Big)$ can then be estimated using the histograms for each of the M stages. A different polar code is constructed for each stage by choosing the frozen bits for a memoryless channel according to the method described in [3], starting with the channel law given by the densities estimated by the Monte-Carlo simulations.

We now explain how interleaved polar codes can achieve the symmetric information rate of the magnetic recording channel when the number of interleaves is large. First, consider the use of a windowed APP-detector that determines the APP based on a fixed window of channel observations with side-length ω. For large enough ω and codeword length N, the window-APP detector achieves the performance of the BCJR algorithm [12]. For such a windowed detector, the output log-APP ratios that are sufficiently separated in time are conditionally independent. Therefore, with large enough M, the equivalent sub-channel seen by each stage is memoryless. Once these equivalent sub-channels have been estimated with sufficient accuracy using the Monte-Carlo simulations described above, the methods known for construction and decoding of polar codes for memoryless channels may be used to achieve rates given by (1). The overall rate $R_{\text{av},M}$ has been shown to converge to the SIR of the ISI channel asymptotically [9], which completes the argument.

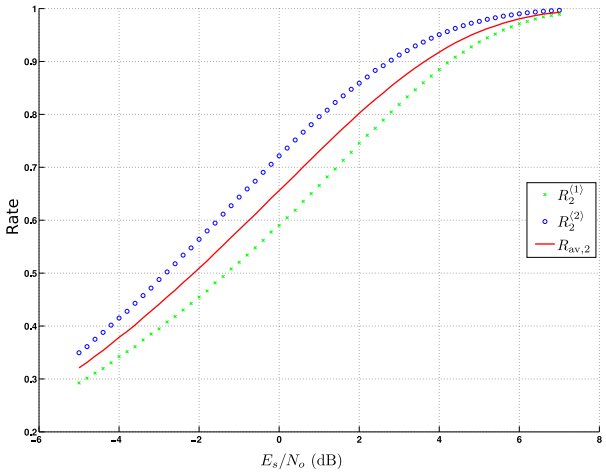

Fig. 7. Achievable rates for 2 stages and average rate for MSD scheme on the dicode channel.

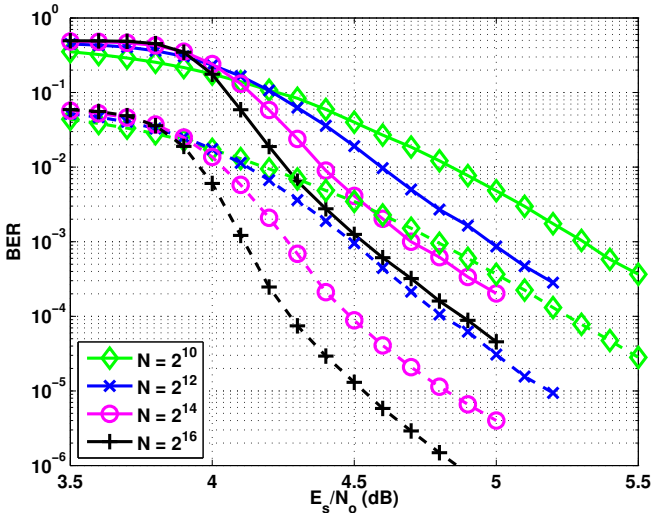

Fig. 8. BER versus SNR for MSD scheme with various blocklengths, N. The dashed and solid lines show the BER for the systematic and non-systematic encodings, respectively.

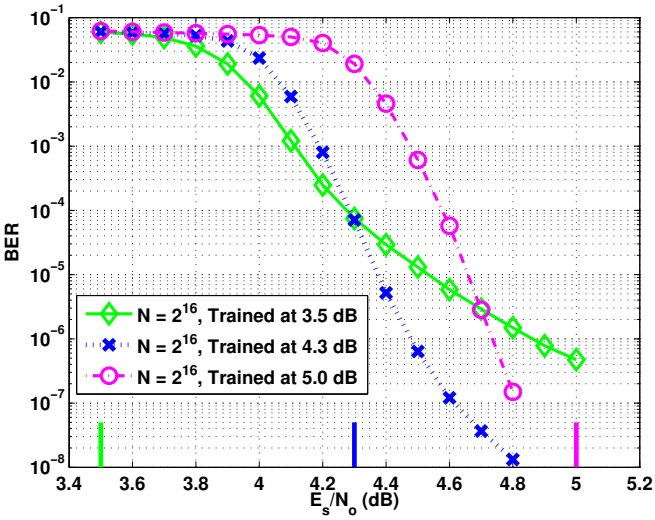

Fig. 9. BER for MSD scheme using SC decoder for polar codes of blocklength $N = 2^{16}$ constructed at different SNR values.

V. Simulation Results

We demonstrate the performance of the proposed MSD scheme on the dicode channel with $h(D) = 1 - D$. First, we simulate the equivalent subchannels for different number of levels to determine the achievable rates given by (1) at various SNR points. The results with $M = 2$ stages are presented in Figure 7. As expected, the rates for successive interleaved stages increase and saturate at the capacity of the AWGN channel with the same noise variance as the ISI channel.

We use the histograms obtained by the Monte-Carlo simulations to estimate the densities $\mathbb{P}\left(L_j^{\langle i\rangle} = l_j^{\langle i\rangle} \,\middle|\, C_j^{\langle i\rangle} = c\right)$ and construct polar codes of the rates given by (1). The choice of frozen bits for these polar codes was done using the algorithm proposed in [3]. The codes were decoded using a SC decoder in the probability domain using double precision floating point numbers. The following results have been obtained by simulations of a 2-stage MSD scheme for the dicode channel. We constructed pairs of polar codes of lengths $N = 2^{10}, 2^{12}, 2^{14}, 2^{16}$ such that $R^{\langle 1\rangle} = 0.85$ and $R^{\langle 2\rangle} = 0.93$. We examined the performance of systematic and non-systematic encoding of polar codes [13]. Figure 8 shows the BER curves for systematic and non-systematic encoding of polar codes. Clearly, the systematic encoding has much better BER performance compared to non-systematic encoding. These results are surprising, given that the SC decoder computes the codeword bits indirectly by first making decisions on the information bits and then encoding the information word back to the codeword. One would expect any decoding errors in the information word to be amplified in the re-encoding process. However, simulation results shown here and those presented earlier in the literature for AWGN channels [13] show that this is not the case. An analytic justification for this phenomenon is still lacking and this represents an important open problem when considering reverse-concatenation with constrained codes, an architecture that is favored in magnetic recording systems.

Since polar codes are constructed for the underlying subchannels which are obtained by Monte-Carlo simulations at a certain SNR of the channel, we also examined how the choice of the design SNR affects the performance of the polar code. Figure 9 shows the BER performance for three pairs of polar codes of the same rate but with different frozen-bit sets which have been determined using simulations at different SNRs. It can be seen that the codes optimized at a slightly lower noise variance perform better. A similar result has been reported for polar codes on memoryless channels in [14]. Also, if the decoders for all polar codes of the same length share the same SC decoding architecture, the knowledge of the operating SNR will optionally allow one to exploit the difference in the performance of these codes of the same rate.

Next we consider the use of the more powerful SC-list decoder. Figure 10 shows the BER improvement as the list

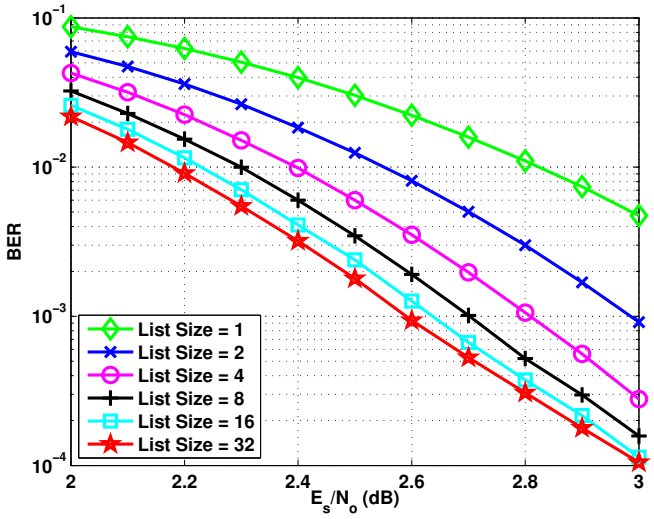

Fig. 10. BER versus SNR for MSD scheme for polar codes with rates $R^{\langle 1 \rangle} = 0.69$, $R^{\langle 2 \rangle} = 0.85$ and blocklength $N = 1024$, decoded with SC-list decoder for various list sizes without CRC-concatenation.

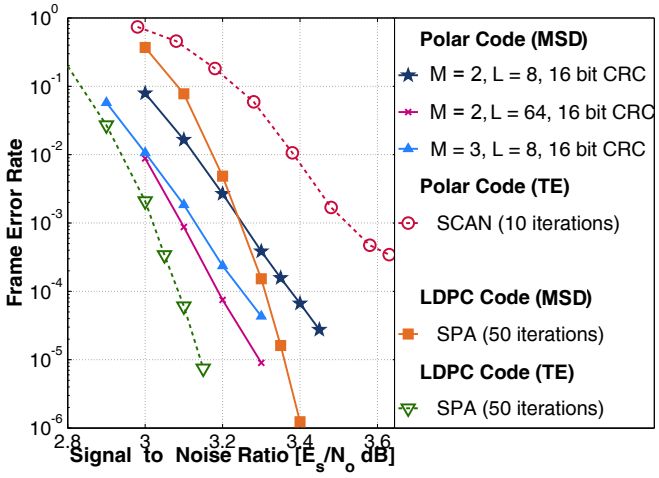

Fig. 11. FER for 2-stage MSD scheme using SC-List decoder for polar codes of blocklength $N = 2^{14}$, with concatenated 16-bit CRC. FER for 2-stage MSD scheme with LDPC codes and for TE scheme with LDPC and polar codes also shown.

size increases. The list decoder with $L = 32$ performs 0.65 dB better than the SC decoder, which is plotted as list size $= 1$. These results match those reported for memoryless channels where at higher SNRs the SC-list decoder quickly approaches the performance of the ML decoder with increasing list sizes.

In order to improve the SC-list decoder, concatenation with an outer CRC code was proposed in [3]. During decoding, the decoder searches for a survivor path which satisfies the CRC. Results in [15] show that this concatenation improves the minimum Hamming distance of the polar codes by eliminating low-weight codewords. We obtained a list of minimum weight codewords for two polar codes of length $N = 2^{14}$ and rates $R^{\langle 1 \rangle} \approx 0.77$ and $R^{\langle 2 \rangle} \approx 0.88$. We found that a 16-bit CRC could eliminate all weight-16 minimum Hamming weight codewords we could find for those codes, using $L \leq 10000$. Figure 11 shows the FER performance of these polar codes. We also plot the performance of the MSD scheme using optimized LDPC codes of equal rates and blocklengths and the performance obtained using the TE scheme, for comparison. The performance of concatenated polar codes can be improved further by increasing the list size as was noted in [15], or by increasing the number of interleaves.

VI. Conclusions

The MSD scheme allows the decoder to assume that the equivalent sub-channels are memoryless, allowing us to use encoding/decoding methods designed for polar codes for binary memoryless channels. For long blocklengths and large number of interleaved stages, polar codes are able to achieve the SIR of the ISI channels. For small number of interleaves, simulation results show that the MSD scheme outperforms the TE scheme when polar codes are used as component codes and are comparable to LDPC codes with the TE scheme.

Acknowledgements

This research was supported in part by NSF Grants CCF-116739 and CCF-1405119, the Center for Magnetic Recording Research at UC San Diego, and Western Digital Corporation.

References

[1] E. Arikan, "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.

[2] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 28 - Jul. 3, 2009, pp. 1496–1500.

[3] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.

[4] R. Pedarsani, S. H. Hassani, I. Tal, and E. Telatar, "On the construction of polar codes," *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 31 - Aug. 5, 2011, pp. 11–15.

[5] I. Tal and A. Vardy, "List decoding of polar codes," *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 31 – Aug. 5, 2011, pp. 1–5.

[6] E. Arikan, "A performance comparison of polar codes and Reed-Muller codes," *IEEE Comm. Letters*, vol. 12, pp. 447–449, Jun. 2008.

[7] U. U. Fayyaz and J. R. Barry, "Polar codes for partial response channels," *Proc. IEEE Int. Conf. Comm.*, Budapest, Hungary, June 2013.

[8] E. Sasoglu, "Polarization in the presence of memory," *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 31 - Aug. 5, 2011, pp. 189–193.

[9] J. B. Soriaga, H. D. Pfister, and P. H. Siegel, "Determining and approaching achievable rates of binary intersymbol interference channels using multistage decoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1416–1429, Apr. 2007.

[10] T. Souvignier, M. Oberg, P. H. Siegel, R. Swanson, and J. K. Wolf, "Turbo decoding for partial response channels," *IEEE Trans. Comm.*, vol. 48, no. 8, pp. 1297–1308, Aug. 2000.

[11] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.

[12] J. B. Soriaga, "On near-capacity code design for partial-response channels," Ph.D. dissertation, Univ. Calif., San Diego, La Jolla, Mar. 2005.

[13] E. Arikan, "Systematic Polar Coding," *IEEE Comm. Letters*, vol. 15, no. 8, pp. 860–862, Aug. 2011.

[14] M. Mondelli, S. Hassani, and R. Urbanke, "From polar to Reed-Muller codes: A technique to improve the finite-length performance," *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 29 - Jul. 4, 2014, pp. 131–135.

[15] B. Li, H. Shen, and D. Tse, "An adaptive successive cancellation list decoder for polar codes with cyclic redundancy check," *IEEE Comm. Letters*, vol.16, no.12, pp. 2044–2047, Dec. 2012.