Polar Codes for Noncoherent MIMO Signalling

Philip R. Balogun, Ian D. Marsland, Ramy H. Gohary, and Halim Yanikomeroglu.
Department of Electrical and Computer Engineering
Carleton University,
Ottawa, Ontario, Canada. K1S5B6
Emails: reinhard.balogun@carleton.ca

Abstract—Polar codes, ever since their introduction, have been shown to be very effective for various wireless communication channels. This together with their relatively low implementation complexity has made polar codes an attractive coding scheme for wireless communications. On the other hand, within the realm of non-coherent wireless MIMO communication, Grassmannian signalling has been shown to approach the ergodic capacity of frequency-flat block fading channels. In this paper, a novel methodology for designing polar codes that works effectively with Grassmannian signalling and a novel set partitioning algorithm for Grassmannian constellations are proposed. We compare the error rate performance of our design with that of existing schemes and show that a gain of over 1 dB over the previously known best technique, which is based on turbo codes, is possible, at much lower decoding complexity.

I. INTRODUCTION

Coherent communication, in which perfect channel state information (CSI) is required at the receiver, dominates current cellular technologies. In order to obtain this CSI, training based estimation methods using pilot or reference signals are used. Future technologies requiring signal detection in multiple-input multiple-output (MIMO) and/or fast fading scenarios would significantly increase the receiver complexity and signalling overhead of such coherent systems. As a result, the noncoherent systems which do not require any channel estimation are therefore attractive as alternatives in such scenarios. The challenge then becomes finding a low-complexity, rate-efficient method of enabling noncoherent communication over MIMO channels.

A possible candidate for noncoherent MIMO communication specifically designed for block-fading channels, are codebooks of unitary matrices that are isotropically distributed on the (compact) Grassmann manifold. These codebooks exploit the MIMO channel characteristics and consider orthogonal subspaces to differentiate the transmitted symbols at the receiver [1].

On the other hand, polar codes are a recent invention by Erdal Arikan that are a class of error-correcting codes with the proven ability to achieve the capacity of binary input, memoryless output symmetric channels. In addition, they provide relatively low complexity encoding and decoding compared with other coding techniques such as turbo codes or LDPC codes.

In the literature, Grassmannian constellations have been combined in a BICM fashion with turbo codes [2]. It was shown that at a high data rate, the scheme was able to outperform training based methods. Further work in this area [3] showed that by improving the labelling for the Grassmannian constellations, the system was able to outperform even training based golden codes.

In this paper, we propose a novel methodology for designing multilevel polar codes that work effectively with Grassmannian signalling and is also applicable to arbitrary, not necessarily structured, multidimensional signalling schemes.

II. SYSTEM MODEL

A. Transmitter

As stated previously, polar codes have the ability to achieve the capacity of binary input memoryless output symmetric channels. These codes are able to achieve this by using an effect known as channel polarization where channels are transformed into good and bad ones (see Fig. 1). By recursively applying such polarization transformation over the resulting channels, the reliabilities of the synthesized channels will show significant difference: the "good ones bet better and the bad get worse" [4]. The channels get more distinctly polarized as the code length is increased and the good channels can be chosen to transmit information bits over while the others are frozen (set to a zero).
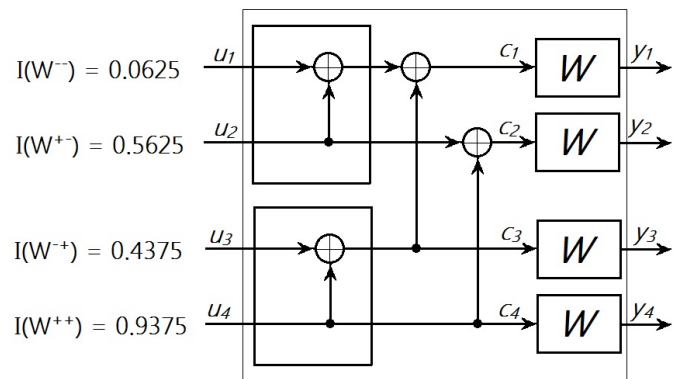

Fig. 1. Sample 4 bit channel polar encoder for a binary erasure channel with capacities shown. Erasure probability = 0.5.

The MLC approach developed by Imai *et al* in [5] and shown in Fig. 2, involves protecting each address bit $B_{i,l}$ of
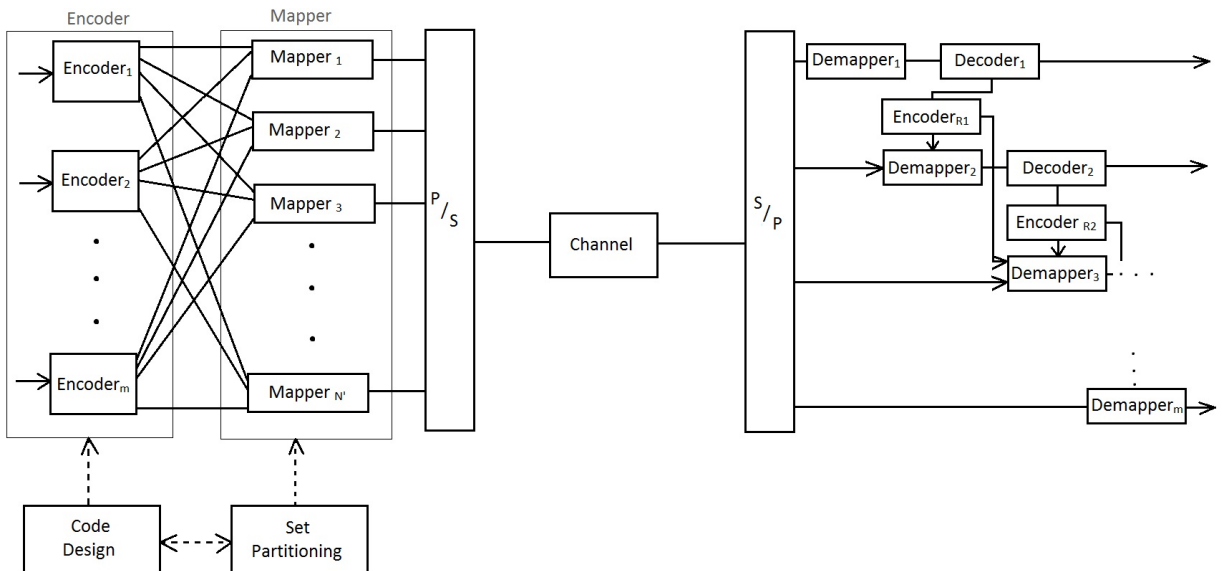
Fig. 2. System set-up.

the signal point by a single binary code at level l [6]. That is, for a signal constellation with M points and $m = \log_2 M$ bits per signal point, there are m encoders – one for each bit level. Each component code is designed to complement the reliability of the bit levels which it encodes, therefore a suitable mapping to the constellation is required to ensure the variance in bit level reliabilities. It was further shown in [6] that the set-partitioning (SP) label for a constellation used in conjunction with multilevel codes is the best technique because SP labelling leads to large bit level variances compared to other types of labelling schemes such as Gray labelling. Each of the m bit levels can therefore be encoded such that the code bits of the first encoder will be mapped to the least reliable bit level while the mth encoder's code bits will be mapped to the most reliable bit level. Each component code has a length N' and therefore the encoders output a total of mN' bits to the mapper. Because each bit level can be thought of as its own channel with a certain capacity, the component code rates are chosen such that the high capacity channels (channels with few error rates) use a higher rate code than those with low capacity. Each component code has a code rate R_i that is chosen such that the overall rate is $R = \frac{1}{m}\sum_{i=1}^{m} R_i$ [7].

B. Receiver

It was shown in [1], [8] that at high SNRs, the ergodic capacity of the channel can be achieved by codebooks of unitary matrices that are isotropically distributed on the (compact) Grassmann manifold. It was also shown that a system with N_t transmit and N_r receive antennas and a coherence time of T will be optimised for high SNR capacity when

$$T \geq \min\{N_t, N_r\} + N_r, \tag{1}$$

$$N_t = \min\left\{\left\lfloor \frac{T}{2} \right\rfloor, N_r\right\}. \tag{2}$$

The received signal is given as

$$\mathbf{Y} = \mathbf{XH} + \sqrt{\frac{N_t}{\rho T}}\mathbf{V}, \tag{3}$$

where $\mathbf{H}$ is an $N_t \times N_r$ channel matrix which has a complex Gaussian distribution, $\mathcal{CN}(0,1)$, and changes independently every block of T channel uses. $\mathbf{V}$ is a $T \times N_r$ matrix which also has a complex Gaussian distribution, $\mathcal{CN}(0,1)$. The average signal to noise ratio, ρ, is calculated per channel use for N_t transmit antennas. $\mathbf{X}$ is a unitary $T \times N_t$ matrix which corresponds to the transmitted signal and has an energy $E_{\mathbf{X}} = N_t$.

The receiver used in this system uses maximum likelihood detection and functions as follows [8]: When $\mathbf{X}$ is multiplied with $\mathbf{H}$, its basis vectors that span an N_t-dimensional subspace are rotated and scaled within the same sub-space. Adding noise to this introduces perturbations within the subspace. The receiver then has to decide which sub-space was transmitted on, irrespective of its basis. This is done by searching the entire constellation to find which one maximizes $p(Y|X)$ where

$$p(Y|X) = \frac{\exp\left(\frac{-1}{\alpha^2}\operatorname{Tr}\left(\mathbf{Y}^{\mathsf{T}}\left(\mathbf{I}_T - \frac{1}{1+\alpha^2}\mathbf{XX}^{\mathsf{T}}\right)\mathbf{Y}\right)\right)}{(\pi\alpha^2)^{TN_r}(1+\frac{1}{\alpha^2})^{N_t N_r}}$$

$$\propto \exp\left\{\frac{||\mathbf{X}^{\mathsf{T}}\mathbf{Y}||^2}{\alpha^2(1+\alpha^2)}\right\}, \tag{4}$$

where α is $\sqrt{\frac{N_t}{\rho T}}$, $\mathbf{I}_T$ is the $T \times T$ identity matrix, the superscript T denotes conjugate transpose and $||\cdot||^2$ is the Frobenius norm.

Having to search through the entire constellation to find the best candidate introduces a high computational cost to this ML detector. Techniques such as the list based decoder exist

which significantly reduce this technique with little cost to the performance [2], but, these are not considered in this paper.

The demapper used in this system is a soft demapper that calculates the log likelihood ratio (LLR) of each bit given the received signal matrix according to [2]:

$$
\begin{aligned}
\lambda_l &= \ln \frac{Pr\{c_l = 0 | Y, c_1, ..., c_{l-1}\}}{Pr\{c_l = 1 | Y, c_1, ..., c_{l-1}\}} \qquad (5) \\
&= \ln \frac{\sum_{X \in \chi_{l,0}} Pr\{Y|X\}}{\sum_{X \in \chi_{l,1}} Pr\{Y|X\}},
\end{aligned}
$$

where Y is the received signal matrix, the set $\chi_{l,k}$ contains all the matrices in the constellation with $B_{i,l} = k$, and the expression $P\{Y|X\}$ is given in (4). As (5) shows, the de-mapping of each bit level relies on knowledge of the originally transmitted bits of the bit levels before it. When the LLRs for a particular level are calculated and the bits are decoded, the receiver passes those bits back into a polar encoder to obtain an approximation of the bits that were transmitted. These approximate codeword bits are then used in the de-mapping of the next bit level and so on (Fig. 2).

III. Set Partitioning for Grassmannian Constellations

Ungerboeck proposed a set partitioning method [9] for regular constellations, such as PSK or QAM, that involves dividing the signal points into two subsets in such a way that the minimum Euclidean distance between any two points in a subset is greater than the minimum distance in the whole constellation. All the points in one subset are assigned a bit value of 0 in the first bit position, while all the points in the other subset are assigned a 1. Each subset is in turn divided into two subsets, again with increasing minimum distance between points within a subset, and all points within one of the new subsets are assigned a bit value of 0 in the second position, while the points within the other new subset are assigned a 1. This process is repeated until each subset contains only one point, and all points have been assigned a unique value of $m = \log_2 M$ bits, where M is the number of points in the constellation. Forney later provided a formalized algorithm for partitioning constellations when the signal points fall on a regular multidimensional lattice [10], but there are no general-purpose algorithms that work with irregular constellations.

Ungerboeck's method works well for regular constellations using the Euclidean distance, and could also be used with small irregular ones, but implementation becomes problematic for large irregular multidimensional constellations. This is because the method is a visual one that will only work with the Euclidean distance as the metric. We therefore propose a novel, formal algorithm that works with any arbitrary constellation, any distance metric, and has low implementation complexity. This algorithm works in the opposite order from Ungerboeck's, starting with M subsets containing just one point each that are paired together to form $M/2$ subsets of two points, which are in turn combined, and so on, until there is one subset of M points.

Given a constellation $\mathcal{C} = \{X_i | i \in \{1, ..., M\}\}$, the ultimate objective of set partitioning is to create a mapping between sequences of $\log_2 M$ bits and points in $\mathcal{C}$. Let $B_{i,l}$ be the lth bit of the mapping for X_i. Furthermore, let $S_{l,i}$ be the set containing the indices of the constellation points in the ith subset at depth l. Note that $|S_{l,i}| = 2^{l-i}$ for $i \in \{1, ..., M/2^{l-1}\}$ and $l \in \{1, ..., \log_2 M\}$. Starting with $S_{l,i} = i$, the algorithm combines subsets at depth $l - 1$ to produce subsets at depth l, based on the inter-subset distance table, $\mathcal{D}_l(i,j) = \min_{a \in S_{l,i}, b \in S_{l,j}} d(X_a, X_b)$, where $d(X_a, X_b)$ is the distance between constellation points X_a and X_b. That is, $\mathcal{D}_l(i,j)$ contains the distance between the two closest points in subsets i and j at depth l. Whereas for coherent detection the Euclidean distance is normally used, it was shown in [8] that for noncoherent detection of Grassmannian constellation the chordal Frobenius norm is more appropriate. The distance between X_a and X_b is given by

$$
d(X_a, X_b) = \sqrt{2N_t - 2 \operatorname{Tr}\left(\Sigma_{X_a^\mathsf{T} X_b}\right)}, \qquad (6)
$$

where N_t is the number of transmit antennas and $\Sigma_{X_a^\mathsf{T} X_b}$ is a matrix containing the singular value decomposition of $X_a^\mathsf{T} X_b$.

When combining subsets it is desirable to pair subsets that are as far apart as possible. That is, a greedy algorithm could pair subset i with subset $\arg \max_j \mathcal{D}_l(i,j)$. We note, however, that at depth l there will always be a pairing with distance $\Delta_l = \min_i \max_j \mathcal{D}_l(i,j)$. Since the system performance depends mostly on this minimum distance, Δ_l, and making the most greedy choices for each i tends to make Δ_{l+1} smaller, we have found it better to pair subset i with the subset that is closest to i but with distance no less than Δ_l. This will ensure that the minimum distance is still Δ_l, without being needlessly greedy. Furthermore, because of numerical rounding errors in calculating the chordal Frobenius norm, a more robust approach is to include a slight error tolerance and accept pairs with distances greater than $\Delta_l - \epsilon$ for some small ϵ (e.g. $\epsilon = 10^{-3}$). The set partitioning algorithm is then formally given by the following pseudo-code.

Algorithm

- Initialize $S_{1,i} \leftarrow \{i\}, M_1 \leftarrow M, \mathcal{D}_1(i,j) \leftarrow d(X_i, X_j)$
- for $l = 1, 2, ..., \log_2 M$
  - $\Delta_l \leftarrow \min_{i \in \{1, ..., M_l\}} \max_{j \in \{1, ..., M_l\}} D_l(i,j)$
  - for $i = 1, 3, ..., M_l - 1$
    * $j \leftarrow \arg \min_{j \in \{i+1, ..., M_l\}, D_l(i,j) > \Delta_l - \epsilon} D_l(i,j)$
    * $B_{a,l} \leftarrow 0 \quad \forall a \in S_{l,i}$
    * $B_{b,l} \leftarrow 1 \quad \forall b \in S_{l,j}$
    * $S_{l+1,(i+1)/2} \leftarrow 0 \quad S_{l,i} \cup S_{l,j}$
    * swap row j of D_l with row $i + 1$
    * swap column j of D_l with column $i + 1$
  - $M_{l+1} \leftarrow M_{l/2}$
  - $\mathcal{D}_{l+1}(i,j) \leftarrow \min\{\mathcal{D}_l(2i - 1, 2j - 1), \mathcal{D}_l(2i - 1, 2j), \mathcal{D}_l(2i, 2j - 1), \mathcal{D}_l(2i, 2j)\} \quad \forall i, j \in 1, 2, ..., M_{l+1}$

IV. System Design Methodology

Due to the polarization effect that occurs in the polar codes, data should be transmitted over the bit channels that have the best channel capacity while the others are not used. With a simple binary erasure channel, it is easy to calculate the capacities of each bit channel, however with our system this computation is not trivial. Therefore, code design was carried out using MATLAB simulations. The code design component of the system is very similar to the set-up described above where an M point Grassmannian constellation is combined with $m = \log_2 M$ component polar encoders each with length N' (where $mN' = N$). Because the positions of the frozen bit channels are not yet known however, the component codes all operates at a rate $R_i = 1$. The design methodology involves simulating the transmission of a large number of message words through the system at a specified design SNR, and recording the bit positions where the first error occurs. Eventually, with enough simulated bits, the bit error rate for each bit channel will be recorded. However, because our system uses a multilevel design where the de-mapping of each successive level depends on the correct detection of all the previous levels, this method would require simulating an extremely large number of bits in order to record first errors in most reliable bit channels. To avoid this, the decoder in the system was changed to provide correct decision feedback every time a decoding error occurred while still keeping track of the positions in which those errors occurred. Once this was done, the system could then accurately estimate the first-error probability for each bit channel. Depending on what coding rate was required, the channels with the smallest first error probabilities would be chosen to transfer our data bits over. The performance of the system can then be evaluated using normal Monte Carlo simulation. However, as show in Fig. 3, the system performance is heavily dependent on the design SNR. Although the designed code tends to work well at the design SNR, its performance can be quite poor at other SNRs. It is often preferable to design a code with a given rate that achieves a given target FER at the lowest possible SNR. For example, from Fig. 3, to achieve a target FER of 10^{-2}, the design SNR of about 10 dB should be used. Since the optimal design SNR is not known in advance, we use a bisection search to find the optimal design SNR.

V. Simulation Results

To investigate the performance of the proposed system, we use an $M = 4096$ point Grassmannian constellation, designed as described in [8], for a 2×2 MIMO channel with a coherence time of $T = 4$. The performance of the rate $\frac{1}{2}$ multi-level polar code, designed as described in Section IV for a target FER of 10^{-2} with $N = 3072$ (12 component codes each with $N' = 256$ bits), for the Grassmannian constellation with labelling as determined according to the new set partitioning algorithm described in Section III, is shown in Fig. 4. The BER performance of the uncoded Grassmannian constellation, with quasi-gray labelling [3] is also shown, and as expected the
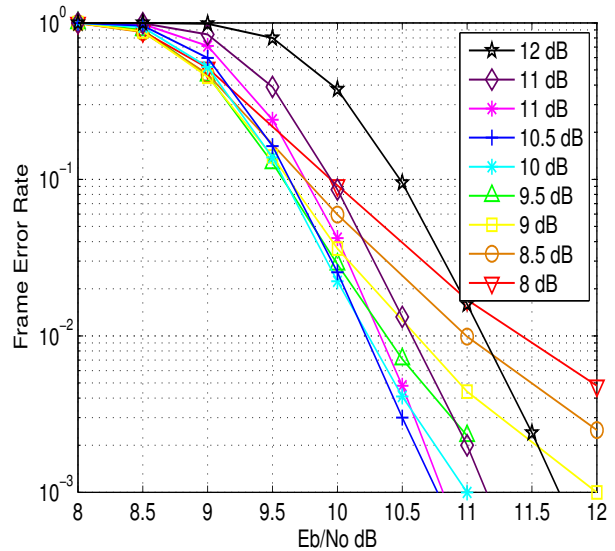

Fig. 3. Effect of design SNR on the FER performance of our system. $N' = 512$, rate $R = 4/5$, $M = 4096$.

coded system shows a significant improvement. For comparison, the performance of a BICM system with a single polar code of length $N = 4096$, designed for the Grassmannian constellation with quasi-Gray labelling is shown. As the figure shows, the new multi-level scheme improves over a single level scheme with a similar value for N by about 4 dB and it improves over the uncoded scheme by about 13 dB at a BER of 10^{-4}.

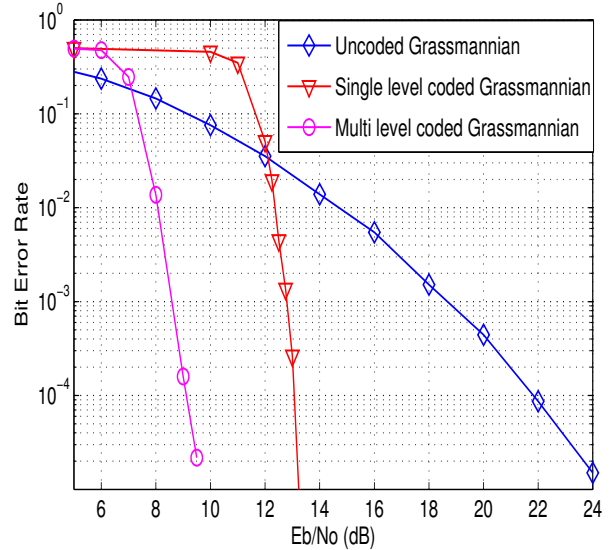
Fig. 4. Comparison of the coded vs uncoded schemes for a 4096 point Grassmannian Constellation. Coded schemes have rate $R = 1/2$. $N = 4096$ for single level code and $N = 3072$ for multilevel code.

As with every coding based system, increasing the frame size improves the bit error rate, but not above the SNR

threshold set by the channel capacity. In Fig. 5 we show the effect of increasing the frame size when the code rate is $4/5$. The figure shows a 4096 point Grassmannian constellation combined with polar codes of varying sub-code lengths ranging from $N' = 8$ to $N' = 8192$ which corresponds to total code lengths of $N = 96$ to $N = 98304$ bits, respectively. As expected, the longer the length of the polar codes, the better the BER performance, however, this improvement is limited by the channel capacity. The SNR limit at capacity was calculated using the expression computed for Grassmannian constellations with equal number of transmit and receive antennas found in [11].

The SNR threshold for the constellation at rate $4/5$ (2.4 bits per channel use) is calculated to be 11.6 dB and this corresponds to an E_b/N_o value of about 7.8 dB. As Fig. 5 shows, with a sub-code length of $N' = 8192$, this design is able to operate within 1.6 dB of the approximate non-coherent ergodic capacity at a BER of 10^{-4}. This is the closest to the channel capacity for a noncoherent system that has been reached as far as we are aware.
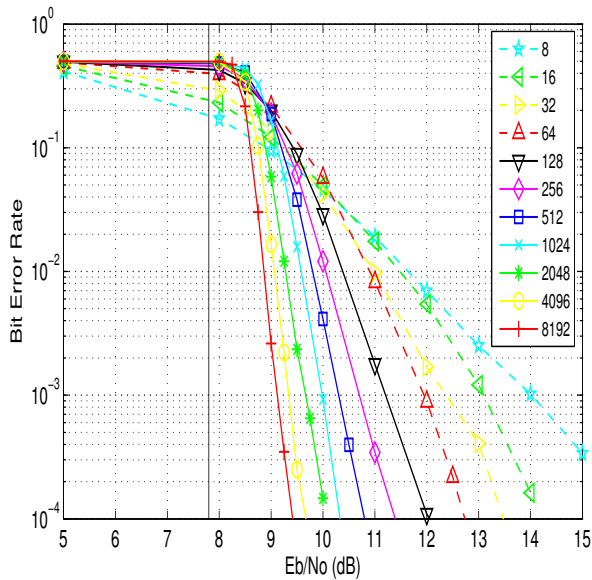

Fig. 5. 4096 point Grassmannian constellation with multilevel polar codes of different sub-code lengths. The overall code rate $R = 4/5$. The SNR threshold is shown at an E_b/N_o of about 7.8 dB.

Thus far, the results seem to indicate that our new system performs quite well. However, it is useful to compare it with other noncoherent techniques that also aim to approach the capacity. Fig. 6 shows the comparison between our polar code design and both turbo and LDPC codes with a 4096 point Grassmannian constellation at a rate of $4/5$. The first two curves give the BER performance of our design using 12 component codes each with $N' = 2048$ and $N' = 4096$, respectively. The next curve shows the performance of a turbo coded system as designed in [3] using a bit interleaved coded modulation with iterative detection and decoding (BICM-IDD), and quasi-gray labelling for the constellation. The last curve gives the performance of a CM system using the stan-dard single level DVB-S2 LDPC code with a Grassmannian constellation. As the figure shows, the polar coding technique performs better than both LDPC and turbo codes for similar code lengths. It is also of note that the polar encoder/decoder has significantly less complexity compared with the other two codes. Therefore, the polar scheme not only provides better error rate performance, but it does so with a much lower complexity.
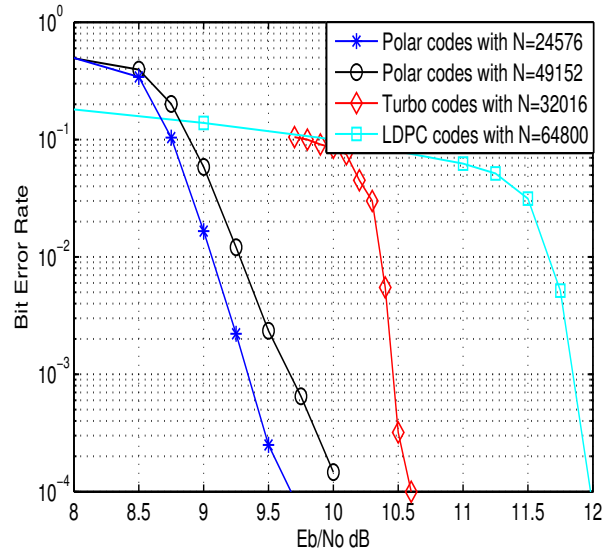

Fig. 6. Comparison of polar codes, turbo codes and LDPC codes with a 4096 point Grassmannian constellation at rate $R = 4/5$.

VI. Conclusion

We have presented a novel combination of polar codes and Grassmannian constellations which to the best of our knowledge has never been attempted before. We were able to adapt multilevel polar coding for Grassmannian constellations and show that this adaptation significantly improves over a basic coded modulation method. As a necessary component to realise these significant gains, a generalised algebraic set partitioning algorithm was developed to create labels for any Grassmannian constellation. This algorithm can also be applied to a wide variety of constellations, even golden codes. The results of the system show that using our code design methodology, we are able to outperform other coding methods such as turbo codes and LDPC codes.

VII. Acknowledgement

References

[1] S. Roger, J. Cabrejas, D. Calabuig, J. F. Monserrat, Y. Fouad, R. H. Gohary, and H. Yanikomeroglu, "Non-coherent MIMO communication for the 5th generation mobile: Overview and practical aspects," *Waves Journal*, no. 6, pp. 5–13, 2014.

[2] L. Szczecinski and A. Alvarado, *Bit-Interleaved Coded Modulation: Fundamentals, Analysis and Design*. Wiley, February 2015.

[3] G. Colman, R. Gohary, M. El-Azizy, T. Willink, and T. Davidson, "Quasi-Gray labelling for Grassmannian constellations," *IEEE Trans. Wireless Commun*, vol. 10, no. 2, pp. 626–636, Feb 2011.

[4] K. Nui, K. Chen, J. Lin, and Q. Zhang, "Polar codes: Primary concepts and practical decoding algorithms," *IEEE Communications Magazine*, vol. 52, pp. 192–203, July 2014.

[5] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp. 371–377, May 1977.

[6] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1361–1391, July 1999.

[7] M. Seidl, A. Schenk, C. Stierstorfer, and J. Huber, "Multilevel polar-coded modulation," in *Proc. IEEE International Symposium on Information Theory*, Istanbul, Turkey, July 2013, pp. 1302–1306.

[8] R. Gohary and T. Davidson, "Non-coherent MIMO communication: Grassmannian constellations and efficient detection," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1176–1205, March 2009.

[9] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets part I: Introduction," *IEEE Communications Magazine*, vol. 25, pp. 5–11, February 1987.

[10] G. D. Forney, "Coset codes – Part I: Introduction and geometrical classification," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 1123–1151, September 1988.

[11] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, February 2002.