On Constructions of Reed-Muller Subcodes

Johannes Van Wonterghem, *Student Member, IEEE,* Joseph J. Boutros, *Senior Member, IEEE,* and Marc Moeneclaey, *Fellow, IEEE*

***Abstract*—In this paper, subcodes constructed from Reed-Muller codes by removal of generator matrix rows are considered. A new greedy algorithm based on the overlap of generator matrix rows is developed. To select the best subcode generated by the greedy algorithm, the number of minimum weight codewords is determined. Computer simulations confirm that the greedy algorithm outperforms the three other construction methods, generating the best codes among all presented subcodes.**

***Index Terms*—Reed-Muller codes, subcodes.**

I. Introduction

Polar codes are proven to achieve capacity [1] for any binary-input memoryless symmetric channel with $O(n \log n)$ complexity encoding and decoding algorithms, where n is the length of the code. Reed-Muller (RM) codes [2] have a construction similar to polar codes. Both can be constructed by selecting rows from the Kronecker product of a same kernel matrix [3]. Also, decoding techniques for polar codes can be applied to RM codes and it has recently been shown that RM codes achieve capacity on the binary erasure channel (BEC) [4], [5] under maximum likelihood (ML) decoding.

How do row selection criteria affect the code performance? The authors of [6] introduced a family of codes that interpolate between RM and polar codes; they have shown that RM codes perform better under ML decoding and polar codes perform better when using a successive cancellation decoder (SCD) [1]. The RM-polar interpolation leads to performance in between ML and SCD using SCD-list [7] or belief propagation decoders [8].

Polar codes are defined by their length $n = 2^m$, an arbitrary code dimension $0 < k < 2^m$ and a channel for which the code is designed. A RM code also has length $n = 2^m$ and is further defined by its degree $0 \leq r \leq m$; we write $\mathcal{R}_{r,m}$ to denote the set of RM codewords. The code dimension is $k = \sum_{i=0}^{r} \binom{m}{i}$. Note that knowledge of the channel is not used while constructing the RM code and that the choice of k is limited to only $m+1$ different values. In [6], this construction was extended to arbitrary k by removing generator matrix rows of an $\mathcal{R}_{r,m}$ code up to the desired value of k. To make a distinction with the classical definition of RM codes, we refer to these codes as RM subcodes.

J. Van Wonterghem is funded by the Research Foundation-Flanders (FWO).

J. Van Wonterghem and M. Moeneclaey are with the Dept. of Telecommunications and Information Processing, Ghent University, 9000 Ghent, Belgium. (e-mail: {johannes.vanwonterghem, marc.moeneclaey}@ugent.be)

J.J. Boutros is with the Dept. of Electrical and Computer Engineering, Texas A&M University at Qatar, Education City, 23874 Doha, Qatar. (e-mail: boutros@tamu.edu)

The authors of [6] did not elaborate on which rows must be removed from the RM generator matrix to get a good subcode. In [9] a procedure for computing the number of minimum weight codewords $N_{\min}$ of RM subcodes is presented and a formula is derived if no more than 3 rows are removed. This allows determining the best subcode for lengths and code dimensions such that no more than 3 rows are removed and such that the search space is manageable. This contribution presents a general heuristic construction method for good RM subcodes, without restrictions on length and dimension. If the procedure for computing $N_{\min}$ of [9] is feasible at the considered length, it allows us to improve the heuristic results.

The codes under consideration are especially interesting for communication systems in the short block-length regime, a topic that has seen an increased interest because of the ongoing 5G standardisation. At long block-lengths, LDPC, turbo and polar codes are still the obvious choice because of their efficient decoding algorithms and good performance. At short block-lengths however, the performance of these codes falls short and alternatives such as BCH or RM codes in conjunction with (near-)ML decoders are viable alternatives.

In section II, four constructions of RM subcodes are presented, including a new greedy algorithm. All subcodes are constructed from the same matrix, so decoders that make use of its recursive structure (e.g., the SCD-list) can be used. As $N_{\min}$ of these subcodes is a dominating multiplicative factor of the error probability at high signal-to-noise ratio, the greedy algorithm is applied in section III to reduce $N_{\min}$ during the subcode construction. Section IV reviews the results of [9] on determining $N_{\min}$. Computer simulation in section V shows that the construction type has a clear impact on the performance of the code, and that the newly proposed greedy construction provides the best results for a given length and dimension of the RM subcode.

II. RM subcodes

Both RM and polar codes can be constructed from the $n \times n$ matrix $G_2^{\otimes m}$ where $n = 2^m$, $G_2 = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$, and the superscript $\otimes m$ represents the m-fold Kronecker product. This definition of $G_2^{\otimes m}$ leads to the following properties [2]:

Property 1: $G_2^{\otimes m}$ has $\binom{m}{i}$ rows of weight 2^i.

Property 2: The rows of $G_2^{\otimes m}$ with Hamming weight w or higher define a code with minimum Hamming distance w.

RM codes, denoted by $\mathcal{R}_{r,m}$, are generated by selecting from $G_2^{\otimes m}$ all rows of weight 2^{m-r} and higher. From Property 1

it follows that the $\mathcal{R}_{r,m}$ code has dimension $k = \sum_{i=0}^{r} \binom{m}{i}$. To construct an $(n = 2^m, k)$ RM subcode where k is not a sum of binomials, rows need to be removed from $\mathcal{R}_{t,m}$ where $t = \min\left\{r : 0 \le r \le m \text{ and } \sum_{i=0}^{r} \binom{m}{i} \ge k\right\}$. In this paper, we limit ourselves to removing some rows of minimum weight from the RM code, i.e. rows that are in $\mathcal{R}_{t,m}$ but not in $\mathcal{R}_{t-1,m}$.

Property 1 shows that there are $\binom{m}{i}$ weight-2^i rows in $G_2^{\otimes m}$. To obtain an $(n = 2^m, k)$ code, we need to choose $s = \sum_{i=0}^{t} \binom{m}{i} - k$ rows to be kept among $\binom{m}{t}$ rows. As an example, consider the $(256, 128)$ RM subcode. It is constructed from the $(256, 163)$ RM code (with degree $r = 4$) by removing 35 of the 70 lowest-weight rows. There are $\binom{70}{35} \approx 10^{20}$ ways of choosing the rows and it is to be expected that not all choices lead to the same error-rate performance. An exhaustive search is not a viable option, so we investigate four strategies for selecting the rows:

1) Construction 1: Sorting the rows of $G_2^{\otimes m}$ in a descending order in terms of the weight (with equal-weight rows maintaining their relative order from $G_2^{\otimes m}$) and selecting the first k rows.
2) Construction 2: Letting $\epsilon \to 0$ in the $(n = 2^m, k)$ polar code construction. Proposition 1 in [6] shows that the resulting code corresponds to a RM code.
3) Construction 3: Randomly choosing s rows among the $\binom{m}{t}$ lowest weight rows to be kept.
4) Construction 4: Greedy algorithm to minimize the overlap between the minimum-weight rows to be kept.

III. Greedy algorithm used in construction 4

Consider an (n, k)-code with $N_{\min}$ codewords of minimum weight $w_{\min}$. Because of the linearity of the RM code, the word error probability $P_{e,\text{word}}$ equals the error probability conditioned on transmitting the all-zero codeword.

An erasure pattern covering a non-zero codeword[1] is a sufficient condition for the ML decoder to make an error on the BEC. The probability of this event, can be bounded from below by the probability that one of the $N_{\min}$ weight-$w_{\min}$ codewords is covered.

Let i be an integer such that $w_{\min} \le i \le n$. There are $\binom{n}{i}$ weight-i erasure patterns, $\binom{n-w_{\min}}{i-w_{\min}}$ of which cover a specific minimum weight codeword. Define $\Psi_{\min}(n, i)$ as the number of erasure patterns of weight i that cover minimum-weight codewords. For $i < w_{\min} + w_{\min}/2$, a weight-$i$ erasure pattern cannot cover multiple minimum-weight codewords, so we have $\Psi_{\min}(n, i) = N_{\min}\binom{n-w_{\min}}{i-w_{\min}}$. The error probability of ML decoding is bounded from below as follows:

$$\begin{aligned} P_{e,\text{word}}(\epsilon) &\ge \sum_{i=w_{\min}}^{w_{\min}+w_{\min}/2-1} \Psi_{\min}(n, i)\epsilon^i(1-\epsilon)^{n-i} \\ &= \sum_{i=w_{\min}}^{w_{\min}+w_{\min}/2-1} N_{\min}\binom{n-w_{\min}}{i-w_{\min}}\epsilon^i(1-\epsilon)^{n-i}, \end{aligned} \quad (1)$$

[1] An erasure pattern is said to cover a non-zero codeword when all 1s of the codeword are erased.

where ϵ denotes the erasure probability. For small ϵ, erasure events covering minimum weight codewords dominate the word error probability and the lower bound is tight enough.

On the BI-AWGN, an approximation of the word error probability can be determined for high signal-to-noise ratio [8]:

$$P_{e,\text{word}}\left(\frac{E_b}{N_0}\right) \approx N_{\min} Q\left(\sqrt{2\frac{k}{n}\frac{E_b}{N_0}w_{\min}}\right), \quad (2)$$

where $Q(.)$ is the Gaussian tail function, and $\frac{E_b}{N_0}$ is the ratio of energy per information bit to the one-sided noise power spectral density.

Because of Property 2, subcodes of the same length and dimension will have the same minimum distance. Equations 1 and 2 hence show that the main difference in ML performance will be due to a difference in $N_{\min}$.

Let g_1 and g_2 be two rows of $G_2^{\otimes m}$ of weight $w_{\min}$. Define the overlap of g_1 and g_2 as $\rho(g_1, g_2) = \sum_{i=1}^{n} g_{1,i}g_{2,i}$, where addition and multiplication are made in $\mathbb{Z}$. If $\rho(g_1, g_2) = w_{\min}/2$, the sum of g_1 and g_2 will generate a codeword of weight $w_{\min}$. For $\rho(g_1, g_2) < w_{\min}/2$, the sum will generate a codeword of weight $> w_{\min}$. Finally, $\rho(g_1, g_2) > w_{\min}/2$ is not possible because of Property 2. If we want to minimize $N_{\min}$, it is hence advised to minimize the overlap between the minimum-weight rows of the generator matrix. This observation motivates Algorithm 1 for the selection of minimum-weight rows. Matrices G_{extra} and T are internal variables of the algorithm.

Algorithm 1 Greedy algorithm of construction 4

Require: $m > 0$ and $0 < k \le 2^m$
1: Construct $G_2^{\otimes m}$.
2: Determine $t = \min\left\{r : 0 \le r \le m \text{ and } \sum_{i=0}^{r} \binom{m}{i} \ge k\right\}$.
3: Fill G_{t-1} with rows of weight up to 2^{m-t+1}.
4: Put 1 row of weight 2^{m-t} in G_{extra} and collect the other rows of weight 2^{m-t} in T.
5: **while** number of rows in $G_{\text{extra}} < s = \binom{m}{t} - k$ **do**
6: Calculate $R(g|G_{\text{extra}}) = \sum_{g' \in G_{\text{extra}}} \rho(g, g')$ for $g \in T$.
7: Move row g with minimum $R(g|G_{\text{extra}})$ to G_{extra}.
8: Collect G_{t-1} and G_{extra} in G and **return** G.

As a result of the recursive structure of $G_2^{\otimes m}$, all choices for the selection of the first row of G_{extra} in line 4 of Algorithm 1 are equivalent. Another result is that multiple g can have the same $R(g|G_{\text{extra}})$ in line 7. When this occurs, an arbitrary g with minimum $R(g|G_{\text{extra}})$ can be chosen (e.g. the first), or a random choice can be made. The latter option results in a non-deterministic output of the algorithm.

IV. Number of minimum-weight codewords

In this section we investigate how to determine $N_{\min}$ for a RM subcode. The knowledge of $N_{\min}$ allows us to make a good choice among a group of subcodes, e.g., those generated by the non-deterministic implementation of the greedy algorithm.

To find $N_{\min}$, we consider the polynomial definition of RM codes. Let $x_1, x_2, ..., x_m$ be m variables. Let $P_{r,m}$ denote the set of binary polynomials with m variables of degree $\leq r$ and let $M_{r,m}$ denote the set of monomials in $P_{r,m}$. For every codeword in $\mathcal{R}_{r,m}$, there exists a unique polynomial in $P_{r,m}$ representing the codeword [2, Ch. 13, §3]. Each polynomial in $P_{r,m}$ can be written as a unique sum of monomials in $M_{r,m}$.

The number of minimum weight codewords in $\mathcal{R}_{r,m}$ is given by [2, Ch. 13, §4]:

$$N_{\min} = 2^r \prod_{i=0}^{m-r-1} \frac{2^{m-i}-1}{2^{m-r-i}-1}. \tag{3}$$

We are interested in subcodes of the $\mathcal{R}_{r,m}$ code. The authors of [9] derived an expression for $N_{\min}(C(\Delta J)$, where $C(\Delta J)$ is the subcode of $\mathcal{R}_{r,m}$ obtained by removing h monomials of degree r from the basis of $\mathcal{R}_{r,m}$. These monomials are represented by $\Delta J = \{\alpha_1, \alpha_2, ..., \alpha_h\}$ with $\alpha_i \in J_r$, where $J_r = \{\{j_1, j_2, ..., j_r\} : 1 \leq j_1 < j_2 < ... < j_r \leq m\}$. The monomial m_{α_i} is the product of the variables $x_{j_1}, x_{j_2}, ..., x_{j_h}$. We reproduce the most important steps of their derivation here.

Define $\Lambda_{r,m}$ as the set of $r \times (m+1)$ binary matrices whose submatrices of the first m columns are of rank r. In [2, Ch. 13, §4] it was shown that the polynomial f represents a codeword of minimum weight 2^{m-r} iff f can be written as:

$$f = p(A) = \prod_{i=1}^{r} \left(a_{i,m+1} + \sum_{j=1}^{m} a_{i,j} x_j \right), \tag{4}$$

where $A \in \Lambda_{r,m}$ and $A = [a_{i,j}]$.

The coefficients of the degree-r monomials in $p(A)$ are found to be $\det(A_\alpha)$, where $\alpha \in J_r$ and where A_α is the submatrix of A constructed from the r columns $\alpha = \{j_1, j_2, ..., j_r\}$.

This allows to determine whether a matrix $A \in \Lambda_{r,m}$ corresponds to a minimum-weight codeword of $C(\Delta J)$. Indeed, a minimum weight codeword of $\mathcal{R}_{r,m}$ can only be a codeword of $C(\Delta J)$ if the monomials corresponding to ΔJ are not present in the polynomial $p(A)$. Or equivalently, iff

$$\det(A_\alpha) = 0, \qquad \text{for } \alpha \in \Delta J. \tag{5}$$

Using the principle of inclusion and exclusion, it can be written that [9, Equ. 5]:

$$N_{\min}\left(C(\Delta J)\right) = 2^r \prod_{i=0}^{m-r-1} \frac{2^{m-i}-1}{2^{m-r-i}-1} + \sum_{s=1}^{h} (-1)^s \sum_{1 \leq i_1 < i_2 < ... < i_s \leq h} \nu(\alpha_{i_1}, \alpha_{i_2}, ..., \alpha_{i_s}). \tag{6}$$

The first term is the number of minimum weight codewords of $\mathcal{R}_{r,m}$ and $\nu(\alpha_{i_1}, \alpha_{i_2}, ..., \alpha_{i_s})$ is the number of codewords in which the monomials $m_{\alpha_{i_1}}, m_{\alpha_{i_2}}, ..., m_{\alpha_{i_s}}$ appear.

In [9, Lemma 1] it was then proven that $\nu(\alpha_{i_1}, \alpha_{i_2}, ..., \alpha_{i_h})$ is equal to the number of $r \times (m+1)$ binary matrices A such that A_{α_1} is the identity matrix and $\det(A_{\alpha_i}) = 1$, for $2 \leq i \leq h$.

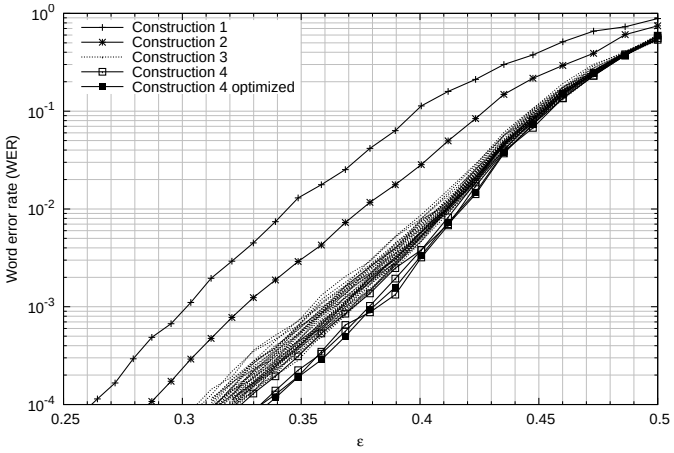

Figure 1. Error rate on the BEC for (256,128) RM subcodes.

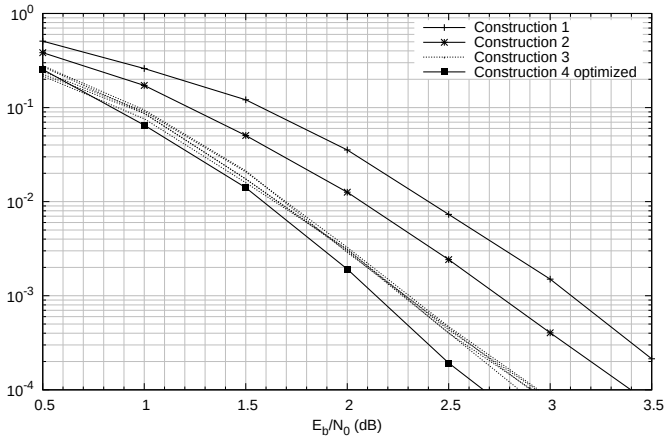

Figure 2. Error rate on the BI-AWGN for (256,128) RM subcodes.

Generating all $2^{(m-r)r}$ binary $r \times m$ matrices $A_{\{1,2,...,m\}}$ with $A_{\{1,2,...,r\}}$ equal to the identity matrix and calculating all determinants from [9, Lemma 1] allows counting $N_{\min}$. The last column of A does not need to be generated because the elements of α_i are in the range $[1, m]$.

V. Performance of subcode constructions

In this section we evaluate the performance of the different constructions of RM subcodes using computer simulation. The word error rates under ML decoding on the BEC and near-ML decoding using the ordered statistics decoding (OSD) algorithm [10] on the BI-AWGN are compared. We refer the reader to [11] for details on the channel models and the decoding algorithms. Note that in practice an SCD-list decoder could be used [7], where the list size allows trading off performance and complexity.

In Fig. 1 and 2 the word error rate on the BEC and BI-AWGN is shown for rate 1/2 RM subcodes of length 256. Multiple subcodes were simulated for construction 3. The non-deterministic version of the greedy algorithm was executed 50 times; in Fig. 1, 'Construction 4' refers to three of these subcodes, whereas 'Construction 4 optimized' refers to the one with the lowest $N_{\min}$. It can be concluded from Fig. 1 and 2

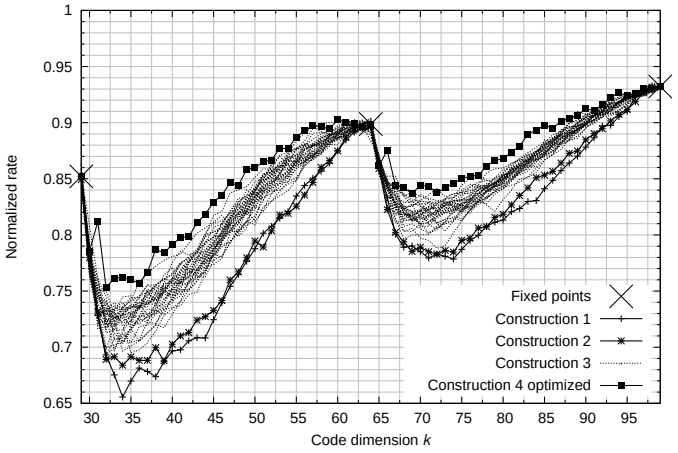

Figure 3. Normalized rates versus code dimension k for length-128 RM subcodes on the BEC, target word error rate of 10^{-3}.

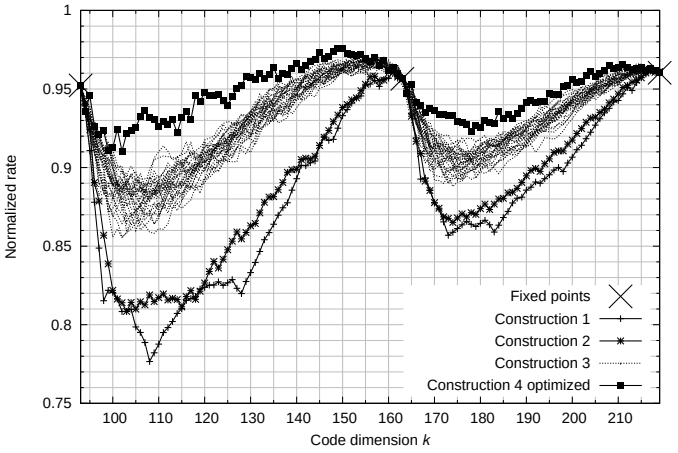

Figure 4. Normalized rates versus code dimension k for length-256 RM subcodes on the BEC, target word error rate of 10^{-3}.

that constructions 1 and 2 are significantly outperformed by constructions 3 and 4. Construction 4 generates subcodes with a performance comparable to the best simulated construction 3 subcodes, or slightly better. Selecting the construction-4 subcode with the lowest number of minimum-weight codewords leads to the best performance.

The normalized rate for the BEC of a given length-n code with $M = 2^k$ codewords was defined in Equ. 299 from [12] as:

$$R_{\text{norm}}(P_{e,\text{word}}) = \frac{\log M}{\log M^*(n, P_{e,\text{word}}, \epsilon_{\min}(P_{e,\text{word}}))}, \qquad (7)$$

where $\epsilon_{\min}(P_{e,\text{word}})$ is the smallest ϵ at which the code still admits decoding below $P_{e,\text{word}}$. For a given n, the maximum achievable code dimension $\log M^*$ on the BEC with given ϵ yielding an error probability below $P_{e,\text{word}}$ is given by Theorem 53 from [12]. The closer $R_{\text{norm}}(P_{e,\text{word}})$ is to 1, the closer the code is to the optimal performance.

Figures 3 and 4 show the normalized rate of, respectively, length-128 and length-256 RM subcodes on the BEC. We indicate the *fixed points*; these are dimensions for which a RM code exists, i.e., these dimensions are sums of binomials. We observe that the performance curves of all constructions intersect at the fixed points. No rows need to be removed to achieve these dimensions, such that all subcodes are equal to the original mother code. The further k moves from the fixed points, the more options there are when constructing the subcodes and the larger the influence of a good construction can be. Similarly, the difference in normalized rate between subcode constructions is higher for the subcodes in Fig. 4 than for those in Fig. 3. Note that, at some values of k, constructions 3 & 4 of length 256 reach a normalized rate closer to 1 than the mother RM code.

Finally, we remark that we also simulated the performance of subcodes generated using a greedy algorithm where $N_{\min}$ is directly used instead of the average correlation. Its high computational complexity makes this construction method impractical, but it demonstrates that the average correlation is indeed a good proxy for $N_{\min}$, as it leads to a subcode with comparable performance without the high computational cost.

VI. Conclusion

We presented four constructions to generate RM subcodes with dimensions that are not sums of binomials. A new greedy algorithm for constructing RM subcodes was proposed. Minimizing the number of minimum-weight codewords is one of the criteria used by the greedy algorithm to select the best subcode. Simulation results under ML decoding on the BEC and near-ML decoding on the BI-AWGN channel were shown. It was demonstrated that the greedy algorithm outperforms the other presented constructions; the performance difference increases when the subcode dimension moves further from a sum of binomials, or when the blocklength increases.

References

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

[2] F. J. MacWilliams and N. J. A. Sloane, "The Theory of Error-Correcting Codes," Amsterdam, The Netherlands: North-Holland, 1977.

[3] E. Arıkan, "A performance comparison of polar codes and Reed-Muller codes," *IEEE Commun. Lett.*, vol. 12, no. 6, pp. 447-449, June 2008.

[4] E. Abbe, A. Shpilka, and A. Wigderson, "Reed-Muller Codes for Random Erasures and Errors," *in IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5229-5252, Oct. 2015.

[5] S. Kudekar et al, "Reed-Muller Codes Achieve Capacity on Erasure Channels," arXiv:1601.04689 [cs.IT], Jan. 2016.

[6] M. Mondelli, S. H. Hassani, and R. Urbanke, "From Polar to Reed-Muller Codes: a technique to improve the finite-length performance," arXiv:1401.3127 [cs.IT], Sept. 2014.

[7] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE International Symposium on Inform. Theory*, St. Petersburg, Russia, pp. 1-5, 2011.

[8] J. G. Proakis and M. Salehi, "Digital Communications," 5th Ed. New York: McGraw-Hill Education, 2007.

[9] H. Tokushige, T. Takata, and T. Kasami, "On the Number of Minimum Weight Codewords of Subcodes of Reed-Muller Codes," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E81-A, no. 10, pp. 1990-1997, Oct. 1998.

[10] M. P. C. Fossorier and S. Lin, "Soft-decision decoding of linear block codes based on ordered statistics," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1379-1396, Sep. 1995.

[11] J. Van Wonterghem, A. Alloum, J. J. Boutros, and M. Moeneclaey, "Performance comparison of short-length error-correcting codes," *2016 Symposium on Communications and Vehicular Technologies (SCVT)*, Mons, Belgium, pp. 1-6, Nov. 2016.

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