Abstract—In this paper, polar codes for the $m$-user multiple access channel (MAC) with binary inputs are constructed. It is shown that Arikan’s polarization technique applied individually to each user transforms independent uses of an $m$-user binary input MAC into successive uses of extremal MACs. This transformation has a number of desirable properties: 1) the “uniform sum-rate” of the original MAC is preserved, 2) the extremal MACs have uniform rate regions that are not only polymatroids but matroids, and thus, 3) their uniform sum-rate can be reached by each user transmitting either uncoded or fixed bits; in this sense, they are easy to communicate over. A polar code can then be constructed with an encoding and decoding complexity of $O(n \log n)$ (where $n$ is the block length), a block error probability of $O(\exp(-n^{1/2-\epsilon}))$, and capable of achieving the uniform sum-rate of any binary input MAC with arbitrary many users. Applications of this polar code construction to channels with a finite field input alphabet and to the additive white Gaussian noise channel are also discussed.

Index Terms—Matroid, multiuser communication, multiple access channel (MAC), polar codes, polarization.

I. INTRODUCTION

The polarization technique, introduced by Arikan [3], transforms $n$ independent uses of a noisy binary input channel into single uses of $n$ synthetic binary input channels. The key property of this transformation is that almost all of these synthetic channels are polarized, in the sense that they are either very noisy or almost noiseless (i.e., having a mutual information either close to 0 or to 1). Moreover, this technique preserves the “uniform mutual information”—the mutual information of the channel with the uniform input distribution—that is, the proportion of synthesized channels that are almost noiseless tends to the uniform mutual information. As the very noisy or almost noiseless channels are channels for which it is easy to code, this transformation leads to the following coding scheme: uncoded information bits are sent on the polarized channels that have uniform mutual informations close to 1, and on the remaining channels, bits frozen to predetermined values are transmitted.

In addition to bringing a new perspective on coding, polar codes can be implemented with low computational effort. More precisely, the encoding and decoding complexity of a polar code is $O(n \log n)$. By definition of the uniform mutual information, these codes achieve the capacity of any channel whose capacity achieving input distribution is uniform. The original polar code construction was generalized in [18] for channels with binary input alphabets to channels with alphabets of arbitrary prime cardinality, allowing polar codes to approach the capacity of any discrete memoryless channel.

In this paper, we show how the polarization technique can be extended to a multiuser problem, namely, the multiple access channel (MAC). One interesting aspect of this extension is that, as opposed to the single-user setting where a single mutual information characterizes an achievable rate, there is in an MAC setting a collection of mutual informations that characterize an achievable rate region. Hence, the terminology “polarized” may need to be revised in an MAC setting, as there may be more than two “polarized MACs.” Indeed, for a two-user binary input MAC, by applying Arikan’s construction to each user’s input separately, Şaşoğlu et al. [19] shows that $n$ independent uses of a two-user MAC are converted into $n$ successive uses of five possible “extremal two-user MACs.” These two-user extremal MACs are the following: 1) each user sees a pure noise channel, 2) one of the users sees a pure noisy channel and the other sees a noiseless channel, 3) both users see a noiseless channel, 4) a pure contention channel: a channel whose uniform rate region is the triangle with vertices $(0,0), (0,1), (1,0)$. Note that for this channel, if any of the two users communicate at zero rate, the other user sees a noiseless channel. Moreover, Şaşoğlu et al. [19] show that the “uniform sum-rate” of the original MAC is preserved during the polarization process, and that the polarization to the extremal MACs occurs fast enough, so as to ensure the construction of a polar code with vanishing block error probability, achieving uniform sum-rate on binary inputs two-user MACs.

In contrast to [19], here we investigate the polarization of the MAC for an arbitrary number of users. In the two-user case, the extremal MACs are not just MACs for which each user sees either a noiseless or pure noise channel, as there is also the pure contention channel. However, the uniform rate regions of the two-user extremal MACs are all polyhedrons with integer-valued constraints. So, the first interesting aspect of the polarization of the MAC with arbitrary many users is to understand what pattern do extremal MACs follow. We will see that the two-user and three-user cases can be handled in a similar manner, whereas a new phenomenon appears when the number of users reaches 4, and the extremal MACs are no longer in a one-to-one correspondence with the polyhedrons having integer-valued constraints. To characterize the extremal MACs,
we first show that the mutual information function used to provide bounds on the communication rates for the extremal MACs corresponds to a specific type of a function called rank function in matroid theory. This connection is used to show that the extremal MACs are in a one to one correspondence with a family of matroids called binary, and are “equivalent” (in a sense which will be defined later) to linear deterministic MACs. This is then used to conclude the construction of a polar code ensuring reliable communication on binary input MACs for arbitrary values of m (the number of users).

Finally, we discuss two applications resulting from the MAC polar code construction with arbitrary many users described in this paper. The first one is motivated by the idea of proposing a new coding scheme for the additive white Gaussian noise channel. By transmitting the standardized average of m binary inputs which are uniformly distributed (taking into account the power constraint), we transmit a random input which is approximately Gaussian distributed when m is large (using the central limit theorem). This is important to achieve the highest rate on the additive white Gaussian noise (AWGN) channel, since the Gaussian input distribution maximizes the mutual information for this channel. We can then use the polar code construction for an MAC developed in this paper to propose a new coding scheme for the AWGN channel. In the second application, we construct polar codes achieving the uniform rate region and to as the uniform sum-rate. We now consider two independent uses of such an MAC. We define two additional binary random vectors

\[ \bar{X}_1 := (X_1[1], \ldots, X_1[m]), \quad \bar{X}_2 := (X_2[1], \ldots, X_2[m]) \]

where \(X_1[i], X_2[i], \) with \(i \in E_m\), are mutually independent and uniformly distributed binary random variables. We denote by \(Y_1\) and \(Y_2\) the respective outputs of independent uses of the MAC \(P\) when the inputs are \(\bar{X}_1\) and \(\bar{X}_2\)

\[ \bar{X}_1 \rightarrow PY_1, \quad \bar{X}_2 \rightarrow PY_2. \]  

We define two additional binary random vectors

\[ \bar{U}_1 := (U_1[1], \ldots, U_1[m]), \quad \bar{U}_2 := (U_2[1], \ldots, U_2[m]) \]

with mutually independent and uniformly distributed components, and we put \(\bar{X}_1\) and \(\bar{X}_2\) in the following one to one correspondence with \(\bar{U}_1\) and \(\bar{U}_2\):

\[ \bar{X}_1 = \bar{U}_1 + \bar{U}_2, \quad \bar{X}_2 = \bar{U}_2 \]

where the addition in the above is the modulo 2 component wise addition.

**Definition 1:** Let \(P : \mathbb{F}_2^m \rightarrow \mathcal{Y}\) be an \(m\)-user BMAC. We define two new \(m\)-user BMACs, \(P^- : \mathbb{F}_2^m \rightarrow \mathcal{Y}^2\) and \(P^+ : \mathbb{F}_2^m \rightarrow \mathcal{Y}^2 \times \mathbb{F}_2^m\), by

\[ P^-(y_1, y_2, \bar{u}_1) := \sum_{u_2 \in \mathbb{F}_2^m} \frac{1}{2^m} P(y_1 | \bar{u}_1 + \bar{u}_2) P(y_2 | \bar{u}_2) \]

\[ P^+(y_1, y_2, \bar{u}_1, \bar{u}_2) := \frac{1}{2^m} P(y_1 | \bar{u}_1 + \bar{u}_2) P(y_2 | \bar{u}_2) \]

for all \(\bar{u}_1, \bar{u}_2 \in \mathbb{F}_2^m, y_i \in \mathcal{Y}, i = 1,2\).

That is, we have now two \(m\)-user BMACs with extended output alphabets:

\[ \bar{U}_1 \rightarrow P^- (Y_1, Y_2), \quad \bar{U}_2 \rightarrow P^+ (Y_1, Y_2, \bar{U}_1) \]

which also defines \(I[J](P^-)\) and \(I[J](P^+), \forall J \subseteq E_m\).

This construction is the natural extension of the construction for \(m = 1,2\) in [3] and [19]. Here again, we are comparing two independent uses of the same channel \(P\) [cf., (2)] with two successive uses of the channels \(P^-\) and \(P^+\) [cf., (3)]. Note that

\[ I[J](P^+) \leq I[J](P^-) \leq I[J](P^+), \quad \forall J \subseteq E_m. \]

**Definition 2:** Let \(\{B_n\}_{n \geq 1}\) be i.i.d. uniform random variables valued in \(\{+,-\}\). Let the BMAC-valued random process \(\{P_n, n \geq 0\}\) be defined by

\[ P_0 := P, \]

\[ P_n := P_{n-1}^{B_n}, \quad \forall n \geq 1. \]

II. POLARIZATION PROCESS FOR MACS

We consider an \(m\)-user MAC with binary input alphabets (BMAC) and arbitrary output alphabet \(\mathcal{Y}\). The channel is specified by the conditional probabilities

\[ P(y | \bar{x}) \quad \text{for all } y \in \mathcal{Y} \text{ and } \bar{x} = (x[1], \ldots, x[m]) \in \mathbb{F}_2^m. \]

Let \(E_m := \{1, \ldots, m\}\) and let \(X[1], \ldots, X[m]\) be mutually independent and uniformly distributed binary random variables. Let \(\bar{X} := (X[1], \ldots, X[m])\). We denote by \(\cal{X}\) the output of the MAC \(P\) when the input is \(\bar{X}\). For \(J \subseteq E_m\), we define

\[ X[J] := \{x[i] : i \in J\} \]

\[ I[J](P) := I(X[J] ; Y | X[J^c]) \]

where \(J^c\) denotes the complement set of \(J\) in \(E_m\), and

\[ I(P) : 2^{E_m} \rightarrow \mathbb{R} \]

\[ J \mapsto I[J](P) \]

where \(2^{E_m}\) denotes the power set of \(E_m\) and where \(I[\emptyset](P) = 0\). Note that

\[ I(P) :\{R_1, \ldots, R_m\} : 0 \leq \sum_{i \in J} R_i \leq I[J](P), \forall J \subseteq E_m. \]
A. Discussion

When \( m = 1 \), we have \( 2I(P) = I(P^{+}) + I(P^{-}) \), which implies that \( I(P_{n}) \) (which in this case denotes a sequence of scalar random variables and not of functions) is a martingale. This allows us to show that \( I(P_{n}) \) tends to either 0 or 1, and the extremal channels of the single-user polarization scheme are either pure noise or noiseless channels. Moreover, in the polarization of the single-user channel, the extremal channels are synthesized by using a genie-aided decoder. The genie helps the decoder in providing the correct values of the previous decisions when decoding the current channel’s input. In the polar code construction, the genie is simulated by a decoder which decodes the bits successively on the synthetic channels, and uses its previous decisions assuming they are correct. As the block error probability of the genie aided and the standalone decoder are exactly the same, it is sufficient to study the block error probability of the genie-aided decoder. These facts then facilitate the design of a code: bits are frozen on the very noisy channels and uncoded information bits are sent on the almost noiseless channels, recovered then by using a successive decision decoder at the receiver. To show that the block error probability of this coding scheme is small, i.e., that the error caused by the successive decision decoder does not propagate, it is shown that the convergence to the “good” extremal channels is fast enough. When \( m \geq 2 \), several new points need to be investigated. In particular, one needs to check whether \( I[J(P_{n})] \) still has a martingale property for different \( J \)’s. Then, if the convergence of each \( I[J(P_{n})] \) can be proved, one has to examine whether the obtained limiting MACs are also extremal MACs, along the spirit of creating trivial channels to communicate over, as in the single-user polarization. Finally, one needs to ensure that the convergence of these mutual informations is taking place fast enough, so as to ensure a block error probability that tends to zero when successive decision decoding is used.

III. PRELIMINARY RESULTS

Summary: In Section III-A, we show that \( I(P_{n}) \) tends a.s. to a matroid rank function (cf., Definition 5). We then see in Section III-B that the extreme points of a uniform rate region with matroidal constraints can be achieved by each user sending uncoded or frozen bits; in particular, the uniform sum-rate can be achieved by such strategies. We then show in Section IV that for arbitrary \( m \), \( I(P_{n}) \) tends not to an arbitrary matroid rank function but to the rank function of a binary matroid (cf., Definition 6). This is used to show that the convergence to the extremal MACs happens fast enough, which then leads to the main result of this paper, Theorem 7 in Section IV. This theorem states that applying Arikan’s polar transform separately to each user, and using a successive decision decoder can achieve sum-rates arbitrarily close to the uniform sum-rate of an MAC, ensure block error probability that decays roughly like \( 2^{-\sqrt{n}} \) with block length, and operate with computational complexity \( O(n \log n) \).

A. Extremal MACs

**Lemma 1:** \( \{I[J(P_{n})], n \geq 0\} \) is a bounded super-martingale when \( J \nsubseteq E_{m} \) and a bounded martingale when \( J = E_{m} \).

**Proof:** For any \( J \subseteq E_{m} \), \( I[J(P_{n})] \leq m \) and

\[
2I[J(P)] = I(X_{1}^1 J; Y_{1} Y_{2} Y_{1}^1 J^c X_{2}^1 J^c) \\
= I(U_{1}^1 J U_{2}^1 J; Y_{1} Y_{2} U_{1}^1 J^c U_{2}^1 J^c) \\
- I(U_{1}^1 J; Y_{1} Y_{2} U_{1}^1 J^c U_{2}^1 J^c) \\
+ I(U_{2}^1 J; Y_{1} Y_{2} U_{1}^1 J^c U_{2}^1 J^c) \\
> I(U_{1}^1 J; Y_{1} Y_{2} U_{1}^1 J^c) \\
+ I(U_{2}^1 J; Y_{1} Y_{2} U_{1}^1 U_{2}^1 J^c) \\
= I[J(P^{-})] + I[J(P^{+})].
\]

If \( J = E_{m} \), the inequality above is an equality.

Note that the inequality in the above are only due to the bounds on the mutual informations of the \( P \) channel. Because of the equality when \( J = E_{m} \), our construction preserves the uniform sum-rate. An illustration of the uniform rate region of \( P^{+}, P^{-} \), and \( P^{0} \) is given in Fig. 1.

As a corollary of previous Lemma, we have the following result.

**Lemma 2:** The process \( \{I[J(P_{n})], J \subseteq E_{m}\} \) converges a.s., i.e., for each \( J \subseteq E_{m} \), \( \lim_{n \to \infty} I[J(P_{n})] \) exists a.s.

Note that for a fixed \( n \), \( \{I[J(P_{n})], J \subseteq E_{m}\} \) denotes the collection of the \( 2^{n} \) random variables \( I[J(P_{n})] \), for \( J \subseteq E_{m} \). When the convergence takes place (this is an a.s. event), let us define

\[
I_{\infty}[J] := \lim_{n \to \infty} I[J(P_{n})]
\]

and \( I_{\infty} \) to be the function \( J \to I_{\infty}[J] \).

From the previous theorem, \( I_{\infty}[J] \) is a random variable valued in \([0, |J|]\). We will now further characterize these random variables.

**Lemma 3:** For any \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that

\[
I(A_{2}; B_{1} B_{2} A_{1}) - I(A_{2}; B_{2}) < \delta
\]

whenever \( \{A_{1}, A_{2}, B_{1}, B_{2}\} \) are random variables valued in \( F_{2} \times F_{2} \times B \times B \), with \( B \) any set, and

\[
P_{A_{1}, A_{2}, U_{1}, U_{2}}(a_{1}, a_{2}, b_{1}, b_{2}) = \frac{1}{4} Q(b_{1}|a_{1} + a_{2}) Q(b_{2}|a_{2})
\]

for any \( a_{i} \in F_{2} \), \( b_{i} \in B \), \( i = 1, 2 \), where \( Q \) is a binary input \( B \)-output channel.
Proof of Lemma 3: First note that
\[ I(A_2; B_1 B_2 A_1) - I(A_2; B_2) 
= I(A_1 + A_2; B_1 B_2 A_1) - I(A_2; B_2) 
\geq I(A_1 + A_2; B_1 B_2) - I(A_2; B_2) 
= H(A_1 + A_2 | B_1 B_2) - H(A_2 | B_2). \]
Moreover
\[ H(A_1 + A_2 | B_1 B_2) 
= \sum_{b_1, b_2 \in \mathcal{B}} H(p_{A_1 | B_1 = b_1} \ast p_{A_2 | B_2 = b_2}) p_{B_1}(b_1) p_{B_2}(b_2) 
\geq \sum_{b_1, b_2 \in \mathcal{B}} h_2(h_2^{-1}(H(A_1 | B_1 = b_1)) \ast h_2^{-1}(H(A_2 | B_2 = b_2))) 
\cdot p_{B_1}(b_1) p_{B_2}(b_2) \]
where \( h_2 \) is the binary entropy function, \( h_2^{-1} \) is its inverse (valued in \([0, 1/2]\)) and where \( a \ast b = a(1 - b) + (1 - a)b. \)
From the so-called Mrs. Gerbers Lemma [21], the function \( x \rightarrow h_2(h_2^{-1}(x) \ast p_0) \) is strictly convex in \( x \) for any \( p_0 \in (1/2, 1) \). Hence, defining \( \alpha = H(A_1 | B_1) = H(A_2 | B_2) \) and using twice this Lemma, we obtain
\[ H(A_1 + A_2 | B_1 B_2) - H(A_2 | B_2) \geq h_2(\alpha \ast \alpha) - h_2(\alpha). \] (6)
Finally, it is straightforward to check that the function \( (0, 1] \ni \alpha \rightarrow h_2(\alpha \ast \alpha) - h_2(\alpha) \) vanishes only when \( \alpha \in \{0, 1\} \). Since this function is continuous, the desired result follows. ■

This Lemma is used to prove the following.

Lemma 4: For any \( \varepsilon > 0 \) and any \( m \)-user BMAC \( P \), there exists \( \delta > 0 \), such that for any \( J \subseteq E_m \), if \( I[J](P^+) - I[J](P) < \delta \), we have
\[ I[J \setminus i](P) \leq I[J \setminus \{i\} | P \in [0, \varepsilon] \cup (1 - \varepsilon, 1), \forall i \in J \]
where \( I[\emptyset](P) = 0 \).

Proof: Let \( i \in J \). Note that
\[ I[J \setminus i](P) = I[U_2 \setminus i; Y_1 Y_2^C U_2 [J^C]] - I(U_2 \setminus i; Y_2 U_2 [J^C]) = I(U_2 \setminus i; Y_1 Y_2^C U_2 [J^C]) \]
\[ \geq I(U_2 \setminus i; Y_1 U_1 [J^C] Y_2 U_2 [J^C]) - I(U_2 \setminus i; Y_2 U_2 [J^C]) = I(U_2 \setminus i; Y_1 X_1 [J^C] Y_2 X_2 [J^C] U_1 [i]) - I(U_2 \setminus i; Y_2 X_2 [J^C]) \]
Using Lemma 3 with \( A_k = U_k \setminus i, B_k = Y_k X_k [J^C], k = 1, 2 \), and
\[ Q(y, x[J^C][x][i]) = \frac{1}{2^{m-1}} \sum_{x[j] \subseteq x[j] \setminus \{i\}} P(y \tilde{x}) \]
we conclude that we can take \( \delta \) small enough, so that \( I[J^C(P^+) - I[J](P) < \delta \) implies \( I(U_2 \setminus i; Y_2 X_2 [J^C]) \in [0, \varepsilon] \cup (1 - \varepsilon, 1) \). Moreover, we have
\[ I[J \setminus i](P) = I[J \setminus \{i\} | P] + I(U_2 \setminus i; Y_2 X_2 [J^C]). \]

Lemma 5: With probability one, \( I[\infty|J] - I[\infty|J \setminus \{i\}] \in \{0, 1\}, \forall J \subseteq E_m, i \in J \), where \( I[\infty, i] = 0 \).

Proof: From Lemma 2, we have that \( I[J](P_n) \) converges a.s., hence \( \lim_{n \rightarrow \infty} I[J](P_n) = I[J](P) \) a.s. Moreover, by definition of \( P_n \), \( I[J](P_{n+1}) - I[J](P_n) \) is equal to \( I[J^C(P^+) - I[J](P) \) w.p. half and \( I[J](P_n) - I[J](P_{n}) \) w.p. half. Hence, from (5),
\[ I[J](P_{n+1}) - I[J](P_{n}) \geq h_2(I[J^C(P^+) - I[J](P)) \]
But \( I[J](P_{n+1}) - I[J](P_{n}) \) is bounded by \( m \), hence \( \lim_{n \rightarrow \infty} I[J](P_{n+1}) - I[J](P_{n}) = 0 \) and \lim_{n \rightarrow \infty} I[J](P_{n+1}) - I[J](P_{n}) = 0 \) almost surely. Finally, we conclude using Lemma 4. ■

Note that Lemma 5 implies in particular that \( \{I[\infty|J] : J \subseteq E_m\} \) is a.s. a discrete random vector.

Definition 3: We denote by \( A_m \), the support of \( \{I[J] : J \subseteq E_m\} \) (when the convergence of Lemma 5 takes place, i.e., a.s.). This is a subset of \( \{0, \ldots, m\}^m \).

We have already seen that not every element in \( \{0, \ldots, m\}^m \) can belong to \( A_m \). We will now further characterize the set \( A_m \).

Definition 4: A polymatroid is a set \( E_m \), a set, equipped with a function \( f : 2^{E_m} \rightarrow \mathbb{R} \) (where \( 2^{E_m} \) denotes the power set of \( E_m \)), called a rank function, which satisfies
\[ f(\emptyset) = 0 \]
\[ f(J) \leq f(K) \quad \forall J \subseteq K \subseteq E_m \]
\[ f(J \cup K) + f(J \cap K) \leq f(J) + f(K) \quad \forall J, K \subseteq E_m. \]
The following result is provided in [9, Lemma 3.1, p. 42].

Lemma 6: For any MAC and any product distribution on the inputs \( X [E_m] \), we have that \( \rho(S) = I(X | S) : Y [S^C] \) is a rank function on \( E_m \), where we denote by \( Y \) the output of the MAC when the input is \( x \) \( E_m \). Hence, \( (E_m, \rho) \) is a polymatroid.

Therefore, any realization of \( I[J] \) is a rank function and the elements of \( A_m \) are the image of a polymatroid rank function.

Definition 5: A matroid is a polymatroid whose rank function is integer valued and satisfies \( f(J) \leq |J|, \forall J \subseteq E_m \). We denote by \( \text{MAT}_m \) the set of all matroids with ground state \( E_m \). We use the notation \( r_B \) to refer to the rank function of a matroid \( B \).

We will sometimes identify a matroid with its rank function image, in which case, we consider an element of \( \text{MAT}_m \) as a \( 2^m \) dimensional integer-valued vector. We also define a basis of a matroid by the collection of maximal subsets of \( E_m \) for which \( f(J) = J \).

Using Lemma 5 and the defintion of a matroid, we have the following result.

Theorem 1: For every \( m \geq 1 \), \( A_m \subseteq \text{MAT}_m \), i.e., \( I[\infty \setminus \{i\}] \) is a matroid rank function.

We will see that the inclusion is strict for \( m \geq 4 \).
B. Communicating on MACs With Matroidal Regions

We have shown that, when $n$ tends to infinity, the MACs that we create with the polarization construction of Section II are particular MACs: the mutual informations $I_{\infty} \{ J \}$ are a.s. integer valued (and satisfy the other matroid properties). As stated in [5, Th. 22], the vertices (corner points) of a polyhedron defined by a rank function $f$ are the vectors of the following form:

\[
x_{j(1)} = f(A_1)
\]

\[
x_{j(i)} = f(A_i) - f(A_{i-1}) \quad \forall i \in \{2, \ldots, k\}
\]

\[
x_{j(i)} = 0 \quad \forall i \in \{k+1, \ldots, m\}
\]

for some $k \leq m$, $j(1), j(2), \ldots, j(m)$ distinct in $E_m$ and $A_i = \{j(1), j(2), \ldots, j(i)\}$, where the vertices strictly in the positive orthant are given for $k = m$.

Therefore, we have the following corollary.

**Corollary 1:** The uniform rate region defined by an element of $A_m$, has vertices on the hypercube $\{0,1\}^m$. In particular, to communicate at a rate $n$-tuple which is a vertex of such an MAC uniform rate region, each user individually communicates on either a noiseless (uniform mutual information of 1) or pure noise (uniform mutual information of 0) channel.

C. Convergence Speed and Representation of Matroids

Convention: for a given $m$, we write the collection $\{I_{\infty} \{ J \}, J \subseteq E_m\}$ by skipping the empty set (since $I_{\infty} \{ 0 \} = 0$) as follows: when $m = 2$, we order the sequence as $(I_{\infty} \{ 1 \}, I_{\infty} \{ 2 \}, I_{\infty} \{ 1, 2 \})$, and when $m = 3$, as $(I_{\infty} \{ 1 \}, I_{\infty} \{ 2 \}, I_{\infty} \{ 3 \}, I_{\infty} \{ 1, 2 \}, I_{\infty} \{ 1, 3 \}, I_{\infty} \{ 2, 3 \}, I_{\infty} \{ 1, 2, 3 \})$, etc.

In this section, we show that there is a correspondence between the extremal MACs and the linear deterministic MACs, i.e., MACs whose outputs are linear forms of the inputs. This correspondence has been used in [19] to establish that the convergence to the extremal MACs for the two-user case is fast, namely $o(2^{-n})$ for any $\beta < 1/2$, which allows to conclude that the block error probability of the code described in [19] is small. We hence follow the same approach as in [19] to treat the case where the number of users is arbitrary, and proceed here to establish this correspondence. We will see that while the case $m = 3$ is similar to the case $m = 2$, a new difficulty is encountered for $m > 4$. How to use this correspondence in order to show that the convergence to the extremal MACs for the $m$-user case is fast is done in Section IV.

Note that, for $m = 2$, a property of the matroids $\{(0, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 1), (1, 1, 2)\}$ is that we can express any of them as the uniform rate region of a linear deterministic MAC: $(1, 0, 1)$ is in particular the uniform rate region of the MAC whose output is $Y = X[1], (0, 1, 1)$ corresponds to $Y = X[2], (1, 1, 1)$ to $Y = X[1] + X[2]$ and $(1, 1, 2)$ to $Y = (X[1], X[2])$. Indeed, this is related to the fact that any matroid with a two element ground state can be represented in the binary field. Let us introduce the definition of binary matroids.

**Definition 6:** Linear matroids: let $A$ be a $k \times m$ matrix over a field. Let $E_m$ be the index set of the columns in $A$. The rank of $J \subseteq E_m$ is defined by the rank of the sub-matrix with columns indexed by $J$.

**Binary matroids:** A matroid is binary if it is a linear matroid over the binary field. We denote by $BMAT_m$ the set of binary matroids with $m$ elements.

1) Case $m = 3$: $MAT_3$ is given by 16 matroids (taking into account the labeling of the ground set). These matroids are represented in Fig. 2. Moreover, they are all binary representable (there are 16 binary matroids). For example, it is clear that the deterministic MAC whose output is $X[1] + X[2] + X[3]$ has a uniform rate region given by $(1, 1, 1, 1, 1)$. Similarly, all matroids for $m = 3$ correspond to the rate region of a linear deterministic MAC. However, one can also show that any three-user binary MAC with uniform rate region given by a matroid is equivalent to a linear deterministic MAC in the following sense. An MAC with output $Y$ and uniform rate region given by $(1, 1, 1, 1, 1, 1)$ must satisfy $I(X[1] + X[2] + X[3]; Y) = 1$, and similarly for other matroids (with $m = 3$), where the linear forms of inputs which can be recovered from the output are dictated by the binary representation of the matroid.

However, the above claim is not quite sufficient to show that, if $\{I_{\infty} \{ J \}, J \subseteq E_m\}$ tends to $(1, 1, 1, 1, 1, 1)$, we have along this path that $I(\{P^1, P^2, P^3\}^m)$ tends to 1, where $P^{1, 2, 3}$ is the channel with input $X[1] + X[2] + X[3]$ and output $Y$. For this, one can show a stronger version of the claim which says that if an MAC has a uniform rate region “close to” $(1, 1, 1, 1, 1, 1)$, it must be that $I(X[1] + X[2] + X[3]; Y)$ is close to 1. In any case, a similar technique as for the $m = 2$ case lets one show that the convergence to the matroids in $A_3$ must take place fast enough.

2) Case $m = 4$: We have that $MAT_4$ contains 68 matroids. However, there are only 67 binary matroids with ground state 4. Hence, there must be a matroid which does not have a binary representation. This matroid is given by $(1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2)$ (one can easily check that this is not a binary matroid). It is denoted $U_{3,4}$ and is called the uniform matroid of rank 2 with four elements (for which any two-element set is a basis). Of course, that this matroid is not binary, does not imply that an hypothetic convergence to
it must be slow. It means that we will not be able to use the
technique employed for the case $m = 2, 3$.

Luckily, one can show that there is no MAC leading to $U_{2,4}$
and the following holds.

**Lemma 7**: $A_2 \subseteq B\text{MAT}_4 \subseteq \text{MAT}_4$. Hence, the $m = 4$ case
can be treated in a similar manner as the previous cases.

We conclude this section by proving the following result,
which implies Lemma 7.

**Lemma 8**: $U_{2,4}$ cannot be the uniform rate region of any
MAC with four users and binary inputs.

**Proof**: Assume that $U_{2,4}$ is the uniform rate region of an
MAC. We then have

$$I(X[i, j]; Y) = 0 \quad (7)$$

$$I(X[i, j]; Y X[k, l]) = 2 \quad (8)$$

for all $i, j, k, l$ distinct in $\{1, 2, 3, 4\}$.

Let $y_0$ be in the support of $Y$. For $x \in F_2^4$, define

$$P(x | y_0) := W(y_0 x) / \sum_{z \in F_2^4} W(y_0 z).$$

Assume w.l.o.g. that $p_0 := P(0, 0, 0, 0 | y_0) > 0$. From (8), a realization
of $X[k, l]$ which has nonzero probability under conditioning of
$y_0$ must determine the value of $X[i, j]$ (for $i, j, k, l$ distinct).
Since $p_0 := P(0, 0, 0, 0 | y_0) > 0$, we must have in particular
that $P(0, 0, *, *) | y_0 = 0$ for any choice of $*, *$ which is not $0, 0$.
And $P(0, 1, *, y_0) = 0$ for any choice of $*, y_0$ which is not $1, 1$.
Since each of these assignments have at least two zeros). On
the other hand, from (7), each time we freeze two components
in $x$ and sum up $P(x | y_0)$ over the other components, we must
obtain the same sum since $X[i, j]$ is uniform given $y_0$. Hence,
$P(0, 1, 1, 1 | y_0)$ must be equal to $p_0$. However, we have from
(8) that $P(1, 0, *, *) | y_0 = 0$ for any choice of $*, *$ (even for $1, 1$
since we now have $P(0, 1, 1, 1 | y_0) > 0$). At the same time, this
implies that the average of $P(1, 0, *, *) | y_0$ over $*, *$ is zero.
This brings a contradiction, since from (7), this average must
equal to $p_0$. ■

Moreover, a similar argument can be used to prove a stronger
version of Lemma 8 to show that no sequence of MACs can have
a uniform rate region that converges to $U_{2,4}$.

**3) Arbitrary Values of $m$**: We have seen in the previous section
that for $m = 2, 3, 4$, the extremal MACs have uniform rate
regions that are not any matroids but binary matroids. This fact
can be used to show that for $m = 2, 3, 4$, $I(J | P_n), J \subseteq E_m$
must tend fast enough to $I_{\infty}(J | J, J \subseteq E_m)$. The details of this
proof are provided in Section IV; in words, by working with the
linear deterministic representation of the MACs, the problem of showing
that the convergence speed is fast in the MAC setting
becomes a consequence of a result shown in [4] for the
single-user setting. We now show that the correspondence
between extremal MACs and linear deterministic MACs holds for
any value of $m$.

**Definition 7**: A matroid is infirmatic if its rank function can
be expressed as $r(J) = I(X[J], Y X[J^c])$, $J \subseteq E_m$, where
$X[E_m]$ has independent and binary uniformly distributed components,
and $Y$ is the output of a binary input MAC with input $x[E_m]$.

**Theorem 2**: A matroid is infirmatic if and only if it is binary.

The converse of this theorem is easily proved and a proof of
the direct part using the following theorem of Tutte [16] can be
found in [1]. We show a stronger result in Theorem 4 below.

**Theorem 3** [16]: A matroid is binary if and only if it has no
minor that is $U_{2,4}$.

A minor of matroids is a matroid which is either a restriction
or a contraction of the original matroid to a subset of the ground
set. A contraction can be defined as a restriction on the dual
matroid, which is another matroid whose bases are the complement
set of the bases of the original matroid. We refer to [14]
for formal definitions. The characterization by a finite number
of excluded minor for a matroid representable over a finite field
is also known for $GF(3)[15], [20]$ and $GF(4)[8]$.

In the following theorem, we connect extremal MACs to
linear deterministic MACs.

**Theorem 4**: Let $X[E_m]$ have independent and binary uniformly
distributed components. Let $Y$ be the output of an MAC
with input $X[E_m]$ and for which $J \subseteq E_m$. Thus $f(J) = I(X[J], Y X[J^c])$
is integer valued, for any $J \subseteq E_m$. Then, there exists a binary matrix
$A$ such that

$$I(AX[E_m]; Y) = \text{rank} A = f(E_m).$$

Note that the first equality in the theorem, namely
$$I(AX[E_m]; Y) = \text{rank} A,$$

is integer valued, for any $X$ from $Y$ can be deduced. Hence, these two facts can be
summarized by saying that $Y$ is “equivalent” to a linear form of
the inputs (in terms of information).

This theorem was originally proved using matroid theory
notations and we refer to [1] for this proof and other inves-
tigations regarding the connection between matroid theory
and extremal MACs. We provide an alternate proof of this theorem
in the Appendix. Note that Theorem 2 follows from Theorem 4.

One can also show a stronger version of this theorem for MACs
having a uniform rate region which is close to a matroid, this is
provided in Theorem 5 below, whose proof is also given in the
Appendix.

**Theorem 5**: Let $X[E_m]$ have independent and binary uniformly
distributed components. For any $\varepsilon > 0$, there exists $\gamma(\varepsilon)$
with the following properties:

1) $\gamma(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$.

2) whenever $Y$ is the output of an MAC with input $X[E_m]$ and

for which $J: 2^E \rightarrow J \rightarrow I(X[J], Y X[J^c])$ satisfies

$$\max_{J, J^c \subseteq 2^E} d(f(J), \mathcal{I}) < \varepsilon,$$

there exists a binary matrix $A$
such that

$$I(AX[E_m]; Y) - f(E_m) < \gamma(\varepsilon).$$
Theorem 2 says that an extremal MAC must have (with probability one) the same uniform rate region as the one of a linear deterministic MAC, i.e., an MAC whose output is a collection of linear forms of the inputs. However, Theorem 4 says something stronger, namely, that from the output of an extremal MAC, one can recover a collection of linear forms of the inputs and essentially nothing else. In that sense, extremal MACs are equivalent to linear deterministic MACs. This also suggests that we could have started from the beginning by working with the quantities $I(P;J) := \sum_{i \in J} X_i; Y_i$ instead of $I(J;P) = I(X[J];YX[J])$, to analyze the polarization of an MAC. The second measure is the natural one to study an MAC, since it characterizes the rate region. However, we have just shown that it is sufficient to work with the first measure to characterize the uniform rate regions of the polarized MACs. Indeed, one can show that $I((P^J)_{\mathbb{Z}})$ tends either to 0 or 1 and Şaşoğlu [17] has provided a direct argument showing that these measures fully characterize the uniform rate region of the extremal MACs. We use a similar argument for the proof of Theorem 4 given in the Appendix.

D. Comment: Relationship Between Information and Matroid Theories

The process of identifying which matroids can have a rank function derived from an information theoretic measure, such as the entropy, has been investigated in different works, cf., [22] and references therein. In particular, the problem of characterizing the entropic matroids has consequent applications in network information theory and network coding problems as described in [11].

Entropic matroids are defined as follows. Let $E$ be a finite set and $X[E] = \{X_i\}_{i \in E}$ be a random vector with each component valued in a finite alphabet. Let $h(I) := h(X[I])$.

**Theorem 6**: $h(\cdot)$ is a rank function. Hence, $(E, h)$ is a polymatroid.

A (polymatroid) is then called entropic, if its rank function can expressed as the entropy of a certain random vector, as above. A proof of the previous theorem is available in [6], [12]. The work in [6], [10], and [22] has resulted in the complete characterization of entropic matroids for $|E| = 2, 3$. However, the problem is open when $|E| \geq 4$. Note that in our case, where we have been interested in characterizing informatic matroids instead of entropic matroids, we have also faced a different phenomenon when $|E| \geq 4$. Other similar problems have been studied in [13].

IV. Main Result: Polar Codes for MACs

In this section, we describe our polar code construction for the MAC and prove the main theorem of the paper.

Let $n = 2^i$ for some $i \in \mathbb{Z}_+$ and let $G_n = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes i}$ denote the $i$-th Kronecker power of the given matrix. Let $U[k] := (U_1[k], \ldots, U_n[k])$ and $X[k]^n = U[k]^n G_n$, $k \in E_m$.

When $X[m]^n$ is transmitted over $n$ independent uses of $P$ to receive $Y^n$, define for any $i \in \{1, \ldots, n\}$ the channel $P_i : F_2^n \rightarrow Y^n \times F_2^{n(i-1)}$ to be the channel whose inputs and outputs are $U_i[m] \rightarrow Y^n U_i^{-1}[m]$. Let $n \geq 1$ and $\varepsilon_n > 0$. Classify each $P_i$ as either ‘polarized’ or ‘not polarized’ according to the function $I(P_i)$ being valued within $\varepsilon_n$ of $0$ or not. (We will choose an appropriate sequence $\{\varepsilon_n\}$ below. For the moment note only that by Theorem 1, if $\varepsilon_n$ were any fixed constant, the channels $P_i$ are in the “polarized” category except for a vanishing fraction of indices $i$.) For $i$ for which $P_i$ is in the polarized category, set $r_i$, to be the integer within $\varepsilon_n$ of $I(P_i)$, that is to say the output of channel $P_i$ determines $A_i U_i[E_m]$ with high probability.

We now describe what we refer to as the polar encoder and decoder for the MAC. The encoder will be specified via the sets $H_i \subset E_m$, the sets of users sending data on $P_i$. These will be chosen as follows: If $P_i$ is not polarized, $H_i$ is empty. Otherwise, select $r_i$ linearly independent columns of the matrix $A_i$, and put $k_i$ in $H_i$ if and only if the $k$th column is selected. For a user $k_i$, let $G[k]$ be the set of $i$ for which $k \in E_i$. For each user $k$ and $i \notin G[k]$, choose $U_i[k]$ independently and uniformly at random, revealing all these “frozen” choices to user $k$ and also to the decoder. The encoder for user $k$ will transmit uncoded bits on channels included in $G[k]$, on the other channels it will transmit the frozen values.

The decoder operates by successively decoding $U_1[E_m], U_2[E_m], \ldots, U_n[E_m]$. At stage $i$, having already decoded $U^{-1}[E_m]$, (assume correctly, for the moment), it is in possession of $(Y^n, U^{-1}[E_m])$, the output of $P_i$. It can thus determine $A_i U_i[E_m]$ with high probability. Since it knows $U_i[B_i^c]$, it can determine $\sum_{k \in H_i} A_i[k] U_i[k]$, and as $\{A_i[k] : k \in B_i\}$ are linearly independent, it can determine $U_i[B_i^c]$.

Observe that for this decoder to operate as described above, it needs the aid of a genie which provides it with $U^{-1}[E_m]$ at stage $i$ of the decoding. Let $U_i[E_m] = \phi_i(Y^n, U^{-1}[E_m])$ denote the decoding function of such a decoder. Observe now, that if we construct an unaided decoder via $U_i[E_m] = \phi_i(Y^n, U^{-1}[E_m])$ using the same decoding function of the genie-aided decoder, the block error event for this unaided decoder $\tilde{U}^n[E_m] \neq U^n[E_m]$ is the same as the block error event $\tilde{U}^n[E_m] \neq U^n[E_m]$ of the genie-aided decoder. Thus, the block error probability of the unaided decoder $P_e(n)$ is equal to the block error probability of the genie-aided decoder and so can be upper bounded as

$$P_e(n) \leq \sum_i P_e(P_i, A_i U_i) \leq \sum_i P_e(P_i, A_i U_i)$$

where $P_e(P_i, A_i U_i)$ is the probability of error in determining $A_i U_i$ from the output of the channel $P_i$. Note now, that we have to be careful in our choice of $\varepsilon_n$: we need to take $\varepsilon_n$ small enough to ensure that $NP_e(P_i, A_i U_i)$ is small. We will see in the proof of Theorem 9 that channel polarization happens so
rapidly that even with such a more stringent choice of $\varepsilon_n$ the
fraction of non polarized channels vanishes with increasing $n$.
(Indeed, for any $\beta < 1/2$, one can choose $\varepsilon_n = 2^{-n^\beta}$ and still ensure polarization.)

Since $I[E_m](P)$ is preserved through the polarization process
(cf., the equality in (5)), we guarantee that with $\delta_n$ denoting the fraction of unpolarized channels

$$
\text{SUM - Rate}(n) := \frac{1}{n}\sum_{k=1}^{m} G[k] = \frac{1}{n} I[E_m](P) - \delta_n - \varepsilon_n.
$$

Thus if $\varepsilon_n \to 0$ is chosen so that $\delta_n \to 0$, the communication system described previously achieves the uniform sum-rate of the underlying channel. The question as to whether $\varepsilon_n$ can be chosen so that both $\delta_n \to 0$ and the block error probability decays to zero is answered in the affirmative by the following theorem.

**Theorem 7:** For any $m \geq 1$, any binary input MAC $P$ with $m$ users, and any $\beta < 1/2$, there exists an integer $n_0$ and a sequence of codes with polar encoders and decoders described above such that the probability of error for a block length $n$ satisfies

$$
P_e(n) \leq 2^{-n^\beta} \quad \forall n \geq n_0
$$

and

$$
\lim_{n \to \infty} \inf_{\varepsilon_n} \text{SUM - Rate}(n) \geq I[E_m](P).
$$

As for the polar code in the single-user setting [3], the encoding and decoding complexity of this code is $O(n \log n)$.

**Proof of Theorem 7:** Fix $\alpha \in (\beta, 1/2)$, $\varepsilon \in (0, 1/2)$ and $\varepsilon_n = 2^{-n^\beta}$. Let $\lfloor x \rfloor$ denote the closest integer to $x$ and define

$$
D_n := \{i \in \{1, \ldots, n\} : I(P_{(i)})[J] \in \mathbb{I} \pm \varepsilon \text{ for any } J \}

\exists A_i \in \mathbb{F}_2^{n \times m} \text{ with } r_i \left[ \begin{array}{c} \int(I(P_{(i)}[E_m])) \end{array} \right]

I(A_i U_i[E_m]; Y^n U[E_m]|^{-1}) > r_i - \varepsilon_n \}.
$$

For $i \in D_n$, we have $H(A_i U_i[E_m]; Y^n U[E_m]|^{-1}) < \varepsilon_n$, and the output of channel $P_{(i)}$ determines $A_i U_i[E_m]$ with high probability, namely

$$
P_e(P_{(i)}; A_i U_i) \leq \varepsilon_n.
$$

Therefore

$$
P_e(n) \leq \sum_{i \in D_n} P_e(P_{(i)}; A_i U_i)
$$

$$
\leq n \varepsilon_n = o(2^{-n^\beta}).
$$

**Definition 8:** For an $m$-user BMAC $P$ with output alphabet $\mathcal{Y}$ and for $S \subseteq b_m$, we define $P[S]$ to be the single-user binary input output alphabet $\mathcal{Y}$, obtained from $P$ by

$$
P[S](y|x) = \frac{1}{2^{m-1}} \sum_{x E_m \in \mathbb{I}} P(y|x E_m)
$$

for all $y \in \mathcal{Y}, x \in \mathbb{F}_2$. Schematically, if $P : X[E_m] \to Y$, we have $P[S] : \sum_{i \in S} X_i \to Y$.

**Lemma 9:** Let $P_{(i)}$ be the channel defined in (9) and let $P[S]$ be the corresponding single-user channel (cf. Definition 8). We have for any $\varepsilon \in (0, 1)$, $\alpha < 1/2$ and $S \subseteq b_m$

$$
\lim_{i \to \infty} -\frac{1}{n} \left[ \int_{i \in \{1, \ldots, n\}} : I(P_{(i)}[S]) > 1 - \varepsilon \right] I((P_{(i)}[S]) < 1 - 2^{-n^\beta}) = 0.
$$

The proof of this lemma is given as follows. Let

$$
\mathcal{D}_n[S] := \{i \in \{1, \ldots, n\} : I(P_{(i)}[S]) > 1 - \varepsilon_n \}

\mathcal{D}_n[S] := \{i \in \{1, \ldots, n\} : I(P_{(i)}[S]) > 1 - \varepsilon \}.
$$

From Lemma 9, we have that

$$
\max_{S \subseteq b_m} \frac{1}{n} \mathcal{D}_n[S] \to 0.
$$

This implies that

$$
\frac{1}{n} \mathcal{D}_n \to 0
$$

where

$$
\mathcal{D}_n := \{i \in \{1, \ldots, n\} : I(P_{(i)})[J] \in \mathbb{I} \pm \varepsilon \text{ for any } J \}

\exists A_i \in \mathbb{F}_2^{n \times m} \text{ with } r_i \left[ \begin{array}{c} \int(I(P_{(i)}[E_m])) \end{array} \right]

I(A_i U_i[E_m]; Y^n U[E_m]|^{-1}) > r_i - \gamma(\varepsilon) \}
$$

where $\gamma(\varepsilon)$ is as in Theorem 5. (The only difference between $\mathcal{D}$ and $\mathcal{D}$ is in the $\gamma(\varepsilon)$ and $\varepsilon_n$ in the last line.)

Since from Theorem 5

$$
\lim_{i \to \infty} -\frac{1}{n} \mathcal{D}_n = 1
$$

we also have from (16)

$$
\lim_{i \to \infty} -\frac{1}{n} \mathcal{D}_n = 1.
$$

Finally, since the polarization process preserves the sum-rate, we conclude the proof of the theorem.

**Proof of Lemma 9:** Note that

$$
(P[S])^{-} \equiv (P^{-}[S])

(P[S])^{+} \leq (P^{+}[S])
$$

where $\equiv$ means that the two transition probability distributions are the same and where $\preceq$ means that they are degraded in the sense

$$
P_1(y|x) \preceq P_2(y|x) \text{ if } P_1(y|x) = P_2(\phi(y)|x)
$$
for some function $\phi$. Hence, defining the Bhattacharyya parameter of a single-user channel $Q$ with binary input and output alphabet $\mathcal{Y}$ by

$$Z(Q) = \sum_{y \in \mathcal{Y}} \sqrt{Q(y|0)Q(y|1)}$$

we have

$$Z([P^+]_{S}) - Z([P[-1]_{S}]^+) \leq 2Z([P[-1]_{S}]^+ - Z([P[-1]_{S}]^2$$

and the random process $Z_t = Z(\{P_t\}_{S})$ satisfies

$$Z_{t+1} \leq Z_t^2 \text{ if } B_{t+1} = 1$$

$$Z_{t+1} \leq 2Z_t \text{ if } B_{t+1} = 0.$$  

We then conclude by using [4, Th. 3], which shows that a random process which satisfies (17) and (18) satisfies for any $\beta < 1/2$

$$\lim_{t \to \infty} \inf P\{Z_t \leq 2^{-\beta t}\} \geq P\{Z_{\infty} = 0\}.$$  

Let us define $p = P\{Z_{\infty} = 1\}$. By definition of $\{Z_{\infty} = 1\}$, for any $\delta \in (0, 1), \eta > 0$, there exists an integer $k$ such that for any $t \geq k$

$$P\{Z_t \geq \delta\} \geq p - \eta.$$  

Hence

$$\lim_{t \to \infty} \inf P\{Z_t \geq \delta\} \geq P\{Z_{\infty} = 1\}.$$  

Therefore

$$P\{2^{-\beta t} < Z_t < \delta\}$$

$$1 - P\{Z_t < 2^{-\beta t}\} - P\{Z_t \geq \delta\}$$

$$0$$

and we proved that

$$\lim_{t \to \infty} \frac{1}{2^n} \sum_{i \in \{1, \ldots, 2^n\}} Z((P(i))_{S}) \leq \delta$$

$$Z((P(i))_{S}) > 2^{-\beta \delta} \text{ for } 2^{-\beta \delta}.$$  

to conclude the proof of the lemma, we use the fact that for any binary input discrete memoryless channel $Q$, we have $I(Q) + Z(Q) \geq 1$ (this is shown in [3, Proposition 11]), and the fact that $I(Q)^2 + Z(Q)^2 \leq 1$ (this is shown in [3, Proposition 1]).

V. EXTENSIONS

A. Power of Prime Cardinality Inputs

The coding scheme developed in this paper can also be used to construct a polar coding scheme (with a deterministic encoding matrix) over single-user channels with input cardinality equal to a power of 2 (or power of prime). Specifically, let $W$ be a DMC with input alphabet cardinality $2^m$ and arbitrary output alphabet. Consider the virtual MAC $W$ obtained by mapping $m$ independent uniform bits into a uniform symbol on $F_{2^m}$ used as an input for the channel $W$. We have that $W$ is an $m$-user binary MAC and the uniform sum-rate of $W$ is the uniform mutual information of $W$. The MAC polar code construction of this paper then a polar code construction for $W$, where the transformation of the inputs over $F_{2^m}$ is done by using the field addition of $F_{2^m}$. More precisely, from $U^n$ i.i.d. uniform over $F_{2^m}$, one construct $X^n = U^nG_n$, where $G_n$ is as define previously (and as in [3]), but where the addition operation used in the multiplication of $U^n$ and $G_n$ is the one of $F_{2^m}$ (which corresponds to the component wise addition of the representation of $U_i$ in $F_{2^m}$). Note that this provides a deterministic construction and an alternative to the scheme proposed in [3] for power of prime cardinality alphabets which uses a randomized construction. Using the result in this paper, one then obtains that the virtual channels obtained with this polarization construction do not have a uniform mutual information which is close to either 0 or $m$, but which is close to an integer in $\{0, 1, \ldots, m\}$. This allows to define a polar coding scheme where part information about the inputs $U_i$ (corresponding to subsets of the components) can be recovered on each extremal channels. Since the uniform sum-rate of $W$ is preserved through the transformation, the uniform mutual information of $W$ is also preserved, and the resulting polar coding scheme achieves the uniform mutual information of $W$. Since $m$ can be arbitrarily large, this provides an alternative to construct capacity achieving polar coding schemes on arbitrary DMCS (using a large enough input cardinality to approximate a nonuniform input distribution).

Similarly, for an MAC with $n$ users and $q$-ary input alphabets, where $q = 2^k$, we can split each user into $k$ virtual users with binary inputs and use the polar code construction of this paper to achieve the uniform sum-rate. Furthermore, if an $n$-user $q$-ary input MAC requires a certain distribution to achieve the (true) sum-rate, then we can split each user into multiple virtual users with binary inputs to approximate the given distribution and thus achieve the sum capacity of an arbitrary MAC.

This can be further generalized to arbitrary prime powers, i.e., to arbitrary finite fields for the input alphabet, although the power of 2 case may be particularly interesting for complexity considerations. Indeed, in that case, one may use fast Fourier transform-like algorithm to reduce the decoding complexity from $O(q^2 \cdot n \log n)$ (general estimate) to $O(q \log q \cdot n \log n)$, where $q = 2^m$ is the input cardinality.

B. Polar Coding for the AWGN Channel

One can use the results of Section IV to construct capacity-achieving codes for the AWGN channel using a quantization scheme for the input distribution. For example, by transmitting the standardized average of i.i.d. binary random variables, scaled to satisfy the power constraint, the receiver observes

$$Y = \frac{2^{\sqrt{p}}}{\sqrt{m}} \sum_{i=1}^{m} (X_i - 1/2) + Z$$

where $Z$ is Gaussian distributed. We can view this channel as being an $m$-user BMAC, $(X_1, \ldots, X_m) \rightarrow Y$, and the polar code constructed in this paper can be used to communicate over this channel. From the central limit theorem, by taking $m$ arbitrarily large, the input distribution of previous scheme is arbitrarily close to a Gaussian distribution, and hence, this coding scheme can achieve rates arbitrarily close to the AWGN capacity. To ensure that this scheme provides a "low encoding and
decoding complexity code” for the AWGN channel, one has to make further complexity considerations when assuming \( m \) arbitrarily large. First, the decoder must recover an \( m \)-dimensional binary vector over each extremal MAC and the total (maximal) number of hypothesis is \( 2^m \). For this, the decoder can proceed with each of the \( m \) users individually (reducing the problem to \( m \) successive hypothesis tests), by using the marginalized single-user channel between one user and the output, which is an extremal channel in the single-user sense. Also, one maximal independent set of users needs to be identified for each extremal MAC, to know where the information bits should be sent. There is no need to check exponentially many sets for this purpose, since this is achieved in at most \( m \) steps, by using a greedy algorithm that checks the independence of a given set and increases the set by one element at each step (starting with the empty set). Polar coding schemes for the AWGN channel have been further investigated in [2].

VI. DISCUSSION

We have constructed a polar code for the MAC with arbitrarily many users, which preserves the properties (complexity, error probability decay) of the polar code constructions in [3] and [19]. The polarization technique brings an interesting perspective on the MAC problem: by polarizing the MACs for each user separately, we create a collection of extremal MACs which are “trivial” to communicate over, both regarding how to handle noise (noiseless or pure noise) but also regarding how to handle interference (which is, modulo synchronization in the code, removed). We have also shown that the extremal MACs are in a one-to-one correspondence with the linear deterministic MACs, i.e., MACs whose outputs are linear forms of the inputs. The polar code constructed in this paper is shown to achieve only a portion of the dominant face of the MAC region, which is, however, sufficient to achieve the uniform sum-rate on any binary input MAC. There are examples of nonextremal MACs where the polar code described in this paper can achieve rates in the entire uniform rate region, for example, the case for a two-user MAC whose output is \( X_1 + X_2 \) with probability half and \( (X_1, X_2) \) with probability half. In general, this may not be the case.

Appendix

In this section, we prove Theorem 4 and Theorem 5. We first need an auxiliary lemma.

**Lemma 10:** Let \( W \) be a binary MAC with two users. Let \( X[E_2] \) with i.i.d. uniform binary components and let \( Y \) be the output of \( W \) when \( X[E] \) is sent. If \( I(X[1]; Y; X[2]) \), \( I(X[2]; Y; X[1]) \) and \( I(X[1]X[2]; Y) \) have specified integer values, then \( I(X[1]; Y), I(X[2]; Y) \) and \( I(X[1] + X[2]; Y) \) have specified values in \( \{0, 1\} \).

**Proof:** Let

\[
I := I(X[1]; Y; X[2]), I(X[2]; Y; X[1]), I(X[1]X[2]; Y)
\]

\[
J := I(X[1]; Y), I(X[2]; Y), I(X[1] + X[2]; Y).
\]

Note that by the polymatroid property of the mutual information, we have

\[
\sum_{x \in E_2} W(x) < 2 \quad \text{and} \quad \sum_{x \in E_2} W(x) \leq 2 \quad \text{for any } x \in E_2.
\]

Let \( y \in \text{Supp}(Y) \) and for any \( x \in E_2 \) define \( P(x|y) = W(y|x) \sum_{z \in E_2} W(y|z) \) (recall that \( W \) is the MAC with inputs \( X[1], X[2] \) and output \( Y \)). Assume w.l.o.g. that \( p_0 := P(0, 0|y) > 0 \).

1) If \( I = 0, 0, 0 \), we clearly must have \( J = 0, 0, 0 \).

2) If \( I = *, *, 1 \), we have \( I(X[2]; Y; X[1]) = 1 \) and we can determine \( X[2] \) by observing \( X[1] \) and \( Y \), which implies \( P(01|y) = 0 \).

Moreover, since \( I(X[1]; Y) = I(X[1]X[2]; Y) - I(X[2]; Y; X[1]) = 0 \), i.e., \( X[1] \) is independent of \( Y \), we must have that \( \sum_{x[2]} P(x[1]x[2]|y) = 1 \) is uniform, and hence

\[
P(01|y) = 1/2, P(10|y) + P(11|y) = 1/2.
\]

Now, if \( I = 1, 1, 1 \), by a symmetric argument as before, we must have \( P(11|y) = 1/2 \) and hence the input pairs 00 and 11 have each probability half (a similar situation occurs when assuming that \( P(x|y) > 0 \) for \( x \neq (0, 0) \)), and we can only recover \( X[1] + X[2] \) from \( Y \), i.e., \( J = 0, 0, 1 \). If instead \( I = 0 \), we then have \( I(X[2]; Y) = I(X[1]X[2]; Y) - I(X[1]; Y; X[2]) = 1 \) and from a realization of \( Y \) we can determine \( X[2] \), i.e., \( P(10) = 1/2 \) and \( J = 0, 1, 0 \).

3) If \( I = 1, 0, 1 \), by symmetry with the previous case, we have \( J = 1, 0, 0 \).

4) If \( I = 1, 1, 2 \), we can recover all inputs from \( Y \), hence \( J = 1, 1, 1 \).

**Proof of Theorem 4:** Let \( I[S](W) \) be assigned an integer for any \( S \subseteq E_m \). By the chain rule of the mutual information

\[
I(X|E_m; Y) = I(X[S]; Y) + I(X[S^c]; Y|X[S])
\]

and we can determine \( I(X[S]; Y) \) for any \( S \). Since for any \( T \subseteq S \)

\[
I(X[S]; Y) = I(X[T]; Y) + I(X[S - T]; Y|X[T])
\]

we can also determine \( I(X[S]; Y|X[T]) \) for any \( S, T \subseteq E_m \) with \( S \cap T = \emptyset \). Hence, we can determine

\[
I(X[1], X[2]; Y|X[S])
\]

\[
I(X[1]; Y|X[S], X[2])
\]

\[
I(X[2]; Y|X[S], X[1])
\]

and thus using Lemma 10, we can determine

\[
I(X[i] + X[j]; Y)
\]

for any \( i, j \in E_m \).
Assume now that we have determined \( I(\sum_T X[i]; Y X[S]) \) for any \( T \) with \( |T| \leq k \) and \( S \subseteq E_m - T \). Let \( T = \{1, \ldots, k\} \) and let \( S \subseteq \{k + 2, \ldots, m\} \)

\[
I(\sum_T X[i], X[k+1]; Y X[S]) = I(X[k+1]; Y X[S]) + I(\sum_T X[i]; Y X[S] X[k+1])
\]

in particular, we can determine

\[
I(X[k+1]; Y \sum_T X[i], X[S]) = I(\sum_T X[i], X[k+1]; Y X[S]) - I(\sum_T X[i], Y X[S])
\]

and

\[
I(\sum_T X[i], X[k+1]; Y X[S]) = I(\sum_T X[i]; Y X[S] X[k+1])
\]

and using Lemma 10, we can determine

\[
I(\sum_T X[i] + X[k+1]; Y X[S])
\]

for any \( T \subseteq E_m \) with \( |T| = k + 1 \). Hence, inducting this argument, we can determine \( I(\sum_T X[i]; Y) \) for any \( T \subseteq E_m \).

Note that the values of these mutual informations must be consistent, for example, if \( I(X[1]+X[2]; Y) = 1 \) and \( I(X[1]+X[3]; Y) = 1 \), we must have \( I(X[2]+X[3]; Y) = 1 \). In general, if \( I(\sum_{T_1} X[i]; Y) = 1 \) and \( I(\sum_{T_2} X[i]; Y) = 1 \), we must have \( I(\sum_{T_3} X[i]; Y) = 1 \) for \( T_3 \subseteq T_1 \cup T_2 \) (the disjoint union of the two sets). Hence, a given MAC induces a given function \( f(-) \), which in turn induces a given collection of subsets \( T \) for which \( I(\sum_T X[i]; Y) = 1 \). These subsets are closed under \( \cup \) and can be generated by a minimal collection of subsets, which form the rows of a matrix \( A \), which satisfies \( I(AX[E_m]; Y) = \text{rank}(A) \). Note that there are several possible minimal collection of subsets and hence several possible matrices \( A \) (all obtained by one of them and by elementary operations over the rows). Moreover, \( f(J) = \text{rank}(A[J]) \), where \( A[J] \) denotes the columns of \( A \) indexed by \( J \). Hence, \( \text{rank}(A) = I(X[E_m]; Y) \).

For example, if \( m = 3 \), \( f(1) = f(2) = f(3) = 1 \), \( f(1, 2) = f(1, 3) = 2 \), \( f(2, 3) = 1 \) and \( f(1, 2, 3) = 2 \), we know that \( g(S) = I(\sum_S X[i]; Y) \) is determined by \( f \) and it is easy to check that in this case \( g(S) = 1 \) if and only if \( S \subseteq \{1, \{2,3\}, \{1,2,3\}\} \) and hence, \( \{1,2,3\} \) is a minimal subset which leads to \( A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \).

In order to prove the “approximative” version of Theorem 4, i.e., Theorem 5, we need the following lemma which is a corollary of Lemma 33 in [19] (Lemma 33 treats only one the cases assumed in Lemma 11, but it is the only nontrivial case as in the proof of Lemma 10). The proof of Theorem 5 follows then from Lemma 11 and the proof of Theorem 4.

**Lemma 11**: Let \( W \) be a binary MAC with two users. Let \( X_E[2] \) with i.i.d. uniform binary components and let \( Y \) be the output of \( W \) when \( X_E \) is sent. If \( I(X[1]; Y X[2]) \), \( I(X[2]; Y X[1]) \) and \( I(X[1], X[2]; Y) \) have specified integer values within \( \varepsilon \), then \( I(X[1]; Y), I(X[2]; Y) \) and \( I(X[1]+X[2]; Y) \) have specified values outside \( (\gamma(\varepsilon), 1 - \gamma(\varepsilon)) \) with \( \gamma(\varepsilon) \rightarrow \varepsilon \rightarrow 0 \).

**References**


Emmanuel Abbe received his M.S. degree in 2003 from the Mathematics Department, Ecole Polytechnique Fédérale de Lausanne, Switzerland, and his Ph.D. degree in 2008 from the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge. Since 2008, he has been working as a postdoctoral fellow and lecturer in the School of Communication and Computer Sciences, Ecole Polytechnique Fédérale de Lausanne, and as a research affiliate at MIT. His research interests include information and coding theory, communications, and combinatorial probability. He received the CVCI award for his master thesis in Mathematics at the Ecole Polytechnique Fédérale de Lausanne in 2003, and the Foundation Latsis International Prize for his postdoctoral research at EPFL in 2011.

Emre Telatar (S’88–M’91–SM’11–F’12) received the B.Sc. degree in electrical engineering from the Middle East Technical University, Ankara, Turkey, in 1986 and the S.M. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 1988 and 1992, respectively. From 1992 to 1999, he was with the Mathematical Sciences Research Center, AT&T Bell Laboratories, Murray Hill, NJ. Since 1999, he has been a Professor at the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. His research interests are in communication and information theories. Dr. Telatar was the recipient of the IEEE Information Theory Society Paper Award in 2001. He was a Program Co-Chair for the IEEE International Symposium on Information Theory in 2002, and an Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2001 to 2004.