Polar coding schemes for the AWGN channel

Emmanuel Abbe
Faculté Informatique et Communications
Ecole Polytechnique Fédérale de Lausanne
Email: emmanuel.abbe@epfl.ch

Andrew Barron
Statistics Department
Yale University
Email: andrew.barron@yale.edu

***Abstract*—This paper investigates polar coding schemes achieving capacity for the AWGN channel. The approaches using a multiple access channel with a large number of binary-input users and a single-user channel with a large prime-cardinality input are compared with respect to complexity attributes. The problem of finding discrete approximations to the Gaussian input is then investigated, and it is shown that a quantile quantizer achieves a gap to capacity which decreases like 1/q (where q is the number of constellation points), improving on the 1/log(q) decay achieved with a binomial (central limit theorem) quantizer.**

I. INTRODUCTION

Polar codes have many desirable attributes [2]. First, they are linear codes with a low encoding and decoding complexity. In addition, their construction is deterministic for most channels (and close to deterministic in general). Finally, they are the first class of low-complexity codes that are provably capacity achieving on any discrete memoryless channels. Here, we consider them for a non-discrete input channel, namely the AWGN channel. We review first the basic construction for binary input alphabets, and in Section III we discuss extensions to larger input alphabets, with emphasis on the complexity analysis. We then discuss in Section IV different constellation schemes to connect polar codes for large input alphabet channels and the AWGN channel.

1) Polar codes for binary input channels: Let P be a binary-input channel with an arbitrary output. Let n be a power of two, U^n be i.i.d. Bernoulli(1/2), $X^n = U^n G_n$ (over $\mathbb{F}_2$) where $G_n = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]^{\otimes \log_2 n}$, and Y^n be the output of n independent uses of P when X^n is the input. Define the mutual information of the so-called synthesized channels by

$$I(P_i) = I(U_i; Y^n U^{i-1}),$$

for $i = 1, \dots, n$. The key result used in the polar code construction is the 'polarization phenomenon', presented here for the binary input case.

Theorem 1. *For any $\varepsilon > 0$,*

$$\frac{1}{n}|\{i \in [n] : I(P_i) \in [0, \varepsilon) \cup (1-\varepsilon, 1]\}| \to 1.$$

That is, except for a vanishing fraction, most of the synthesized channels are either very noisy or almost perfect. This suggests the following coding scheme.

Polar encoder: For a threshold ε (which will upper bound the probability of error), define the set of 'good' components by

$$\mathcal{G}_\varepsilon = \{i \in [n] : I(P_i) \geq 1 - \varepsilon\}.$$

On these components, transmit uncoded bits. On $\mathcal{G}_\varepsilon^c$, draw the bits uniformly at random and reveal their values to the decoder. In the case where the channel is symmetric, one can equally well freeze these bits deterministically and arbitrarily. Note that the rate of this code is given by $|\mathcal{G}_\varepsilon|/n$ which we know to be $I(P) + o_n(1)$ by Theorem 1 (where $I(P)$ denotes the mutual information with uniform input).

Polar decoder: Being in possession of $Y^n = y^n$ and $U^n[\mathcal{G}_\varepsilon^c]$, the decoder proceeds as follows.

(0) Define $T = \mathcal{G}_\varepsilon$ and $\hat{u}[T] = u[T]$.

(1) Find the smallest component i in T and compute

$$\hat{u}_i = \arg\max_{u \in \mathbb{F}_2} \mathbb{P}\{U_i = u | Y^n = y^n, U^{i-1} = \hat{u}^{i-1}\}. \quad (1)$$

Update $T = T \setminus i$ and $\hat{u}[T] = \hat{u}[T \cup i]$.

(2) Iterate (1) until T is empty.

The probability of error is then given by $\mathbb{P}\{\hat{U}^n \neq U^n\}$.

Theorem 2. *[3] The polar coding scheme defined previously allows to achieve the uniform mutual information*[1] *of any binary-input discrete memoryless channel, with an error probability of $O(2^{-n^\beta})$, $\beta < 1/2$, and an encoding and decoding complexity of $O(n \log n)$.*

If the input distribution maximizing the mutual information of the channel is uniform, the uniform mutual information is the capacity. This is the case if the channel is symmetric.

A. Discrete codes for the AWGN channel

The problem of modulating discrete codes to approach the AWGN channel capacity is an old problem. To goal of this paper is to compare the different approaches that are available *within polar codes*, in particular the use polar codes for large prime alphabets and for MACs. The idea of using a MAC (with binary inputs) to approach the AWGN channel capacity is also well-known, as discussed in [12]. A major part of the existing literature as however focused on the problem of optimally choosing a finite number of constellation points, as discussed in [6], which provides a good survey of this literature, in particular of [5], [7]. Yet, we are interested here in a slightly different problem, namely, in studying the scaling of the gap to capacity when the number of constellation point grows, and when the constellation points must have equal probability (which results from polar codes requirements). Our

[1] the uniform capacity of a channel is the mutual information of the channel when the input has a uniform input distribution

goal is to achieve a fast enough decay, say inverse-linear in the number of constellation points, to cope with the complexity requirements of the polar coding schemes discussed in Section III. There is a much smaller literature studying the scaling of the gap to capacity for such constellations. In [10], it is shown that this gap goes to zero, which is to be expected. In [11], a fast decay is shown but for arbitrary constellations which are not equiprobable. To the best of our knowledge, the problem of finding the scaling of the gap to capacity for equiprobable constellations has not been further studied. We show in Section IV that the scaling is at least inverse-linear, which is sufficient to cope with the complexity requirement of the polar coding schemes of Section III and which significantly improves on the inverse-logarithmic decay of the central limit theorem approach used in [1], [7].

II. Goals and results

Our goal is to design an efficient polar coding scheme for the AWGN channel. We first consider polar coding schemes for large input alphabets, comparing the MAC approach [1] (with many users) and the single-user approach [9] (with large prime-cardinality inputs). We then investigate the problem of constructing good finite constellations for the AWGN channel, which we then combine with the polar code approaches.

The main results of this paper show that:

- The MAC approach with many users shows better complexity attributes than the single-user approach with large prime alphabets
- The constellation initially proposed in [1] using the central limit theorem can be significantly improved upon, with a quantile quantizer whose gap to capacity decay is SNR$/q$, where q is the number of constellation points.

III. Polar codes for large input alphabets

A. Prime alphabets

Let P be a q-ary input discrete memoryless channel, where q is prime. Let U^n be i.i.d. uniform on $\mathbb{F}_q$, $X^n = U^n G_n$ (over $\mathbb{F}_q$) and Y^n be the output of n independent uses of P when X^n is the input. Define the mutual information of the synthesized channels (through the polar transform) by

$$I(P_i) = I(U_i; Y^n U^{i-1}),$$

for $i = 1, \ldots, n$, where the logarithm in the mutual information is computed in base q. Then the conclusion of Theorem 1 holds readily, using the logarithm in base q for the mutual information computations.

Define now the encoder and decoder exactly as in Section I-1, replacing $\mathbb{F}_2$ by $\mathbb{F}_q$.

Theorem 3. *[9] The polar coding scheme defined previously allows to achieve the uniform mutual information of any q-ary input discrete memoryless channel when q is prime, with an error probability of $O(2^{-n^\beta})$, $\beta < 1/2$, and an encoding and decoding complexity of $O(n \log n)$.*

Note that in previous theorem, the dependency in q does not appear in the error probability and complexity estimates. For large q, this may matter. By inspection of [9], one finds that the probability of error scales at most multiplicatively with q, i.e., as $O(q 2^{-n^\beta})$. Regarding the complexity, one needs to inspect the following optimization step in the decoding algorithm

$$\hat{u}_i = \arg\max_{u \in \mathbb{F}_q} \mathbb{P}\{U_i = u | Y^n = y^n, U^{i-1} = \hat{u}^{i-1}\}.$$

As for the binary case [2], the computation of $\mathbb{P}\{U_i = u | Y^n = y^n, U^{i-1} = u^{i-1}\}$ can be done with a divide and conquer algorithm, which computes the synthesized channels $P_i^{(n)}$ for the block length level n by calling the computations of two other synthesized channels $P_j^{(n/2)}$ and $P_k^{(n/2)}$ corresponding to the block length level $n/2$. It is easy to check that for a fixed i, this requires $n \log_2 n$ computations of either the $+$ or $-$ operations of a q-ary input channel P, where

$$P^-(y_1 y_2 | u_1) = \frac{1}{q} \sum_{u_2 \in \mathbb{F}_q} P(y_1 | u_1 + u_2) P(y_2 | u_2)$$

$$P^+(y_1 y_2 u_1 | u_2) = \frac{1}{q} P(y_1 | u_1 + u_2) P(y_2 | u_2).$$

Note that P^- or P^+ require now q^2 computations. Therefore, with the same argument as for the binary case, the total decoding complexity is $O(q^2 n \log_2 n)$.

B. Powers of prime alphabets and MAC polar codes

The results in [2], [9] (presented in previous sections), in particular Theorem 1, do not extend to channels whose input alphabets cardinality are not prime but powers of prime. An alternative proposed in [2] consist in randomizing the polar transformation, i.e., replacing G_n with a properly adapted randomized matrix. Another alternative is yet possible with the same matrix G_n, using the construction of polar codes for the m-user MAC [1]. We now review this framework.

Let W be an m-user binary input MAC (the case of prime input alphabets can be treated in a similar manner, as explained later, powers of two are of particular interest here). Let U^n be i.i.d. uniform over $\mathbb{F}_2^m$ (note that these are vectors of dimension m), $X^n = U^n G_n$ (over $\mathbb{F}_2$) and Y^n be the output of n independent uses of W when X^n is the input. Define the 'mutual information vectors' of the synthesized MACs (through the polar transform) by

$$I(W_j) = \{I(W_j)[S] : S \subseteq [m]\},$$

where

$$I(W_j)[S] = I(U_j[S]; Y^n U^{j-1} U_i[S^c]),$$

for $j = 1, \ldots, n$.

The generalization of Theorem 1 to the MAC setting is as follows.

Theorem 4. *[1]*

$$|\{j \in [n] : I(W_j) \in \{0, 1, \ldots, m\} \pm \varepsilon\}| \to 1.$$

This theorem suggests the following coding scheme.

MAC polar encoder: For a threshold ε (which will lead to an upper bound on the probability of error), define the set of 'good' components by

$$\mathcal{G}_\varepsilon = \cup_{j=1}^n \mathcal{G}_{j,\varepsilon},$$
$$\mathcal{G}_{j,\varepsilon} = \arg\max\{|S| : I(W_j)[S] \geq |S| - \varepsilon,\, S \subseteq [m]\}, \quad (2)$$

and if the maximizer is not unique in (2), pick one arbitrarily (note that we can have $\mathcal{G}_{j,\varepsilon} = \emptyset$). On the components indexed by $\mathcal{G}_\varepsilon$, transmit uncoded bits. On $\mathcal{G}_\varepsilon^c$, draw the bits uniformly at random and reveal the realizations to the decoder. One can also define symmetric MACs for which these bits can frozen deterministically and arbitrarily. Note that the sum-rate of this code is given by $|\mathcal{G}_\varepsilon|/n$ which can be shown to be $I(W)[1,\dots,m] + o_n(1)$, using Theorem 4 (where $I(W)[1,\dots,m]$ denotes the uniform sum-rate).

MAC polar decoder: Being in possession of $Y^n = y^n$ and $U^n[\mathcal{G}_\varepsilon^c] = u^n[\mathcal{G}_\varepsilon^c]$, the decoder proceeds as follows.

(0) Define $S_j = \mathcal{G}_{j,\varepsilon}$ and $\hat{u}[S_j] = u[S_j]$ for all $j = 1,\dots,n$.

(1) For $j = 1,\dots,n$, if $S_j \neq \emptyset$ compute

$$\hat{u}_j[S_j] = \arg\max_{u \in \mathbb{F}_2^{|S_j|}} \quad (3)$$
$$\mathbb{P}\{U_j[S_j] = u | Y^n = y^n, U^{j-1} = \hat{u}^{j-1}, U_j[S_j^c] = \hat{u}_j[S_j^c]\}.$$

The probability of error is then given by $\mathbb{P}\{\hat{U}^n \neq U^n\}$.

Theorem 5. *[1] The polar coding scheme defined previously allows to achieve the uniform sum-mutual information [2] of any m-user binary input discrete memoryless MAC, with an error probability of $O(2^{-n^\beta})$, $\beta < 1/2$, and an encoding and decoding complexity of $O(n \log n)$.*

We can then use this result to construct polar coding schemes for single-user channels having q-ary input alphabets when $q = 2^m$. Indeed, for a 2^m-ary input alphabet channel P, define W to be the m-user binary input MAC given by

$$W(y|x_1,\dots,x_m) = P(y|f(x_1,\dots,x_m))$$

where f is a bijection. Then, the uniform mutual information of P is the uniform sum-mutual information of W and we have the following from Theorem 5

Corollary 1. *The polar coding scheme defined previously allows to achieve the uniform mutual information of any q-ary input channel when q is a power of two, with an error probability of $O(2^{-n^\beta})$, $\beta < 1/2$, and an encoding and decoding complexity of $O(n \log n)$.*

This result extends to powers of primes, although powers of two are of particular interest here.

We now investigate the dependency in q of previous orders. By inspection of [1], one finds that the probability of error scales at most multiplicatively with m. Regarding the complexity, one needs to inspect the following optimization step in the decoding algorithm

$$\hat{u}_j[S_j] = \arg\max_{u \in \mathbb{F}_2^{|S_j|}}$$
$$\mathbb{P}\{U_j[S_j] = u | Y^n = y^n, U^{j-1} = \hat{u}^{j-1}, U_j[S_j^c] = \hat{u}_j[S_j^c]\}.$$

As for the binary case [2], the computation of these probabilities can be done with a divide and conquer algorithm, which computes the MACs $W_j^{(n)}$ by calling the computations of two other synthesized MACs $W_j^{(n/2)}$ and $W_k^{(n/2)}$ corresponding to the $n/2$-block length level. Again, for a fixed j, this requires $n \log_2 n$ computations of either the $+$ or $-$ operations of a m-user binary input MAC W, where

$$W^-(y_1 y_2|u_1) = \frac{1}{q} \sum_{u_2 \in \mathbb{F}_2^m} W(y_1|u_1+u_2) W(y_2|u_2)$$
$$W^+(y_1 y_2 u_1|u_2) = \frac{1}{q} W(y_1|u_1+u_2) W(y_2|u_2).$$

In the single-user q-ary case where q is prime, each of these convolution operations require at most q^2 operations. However, here, we can take advantage of the fact that $q = 2^m$ here and that previous operations are convolution-like. Hence, using an algorithm a la FFT, we can bring down the complexity of these operations to $O(q \log_2 q)$ instead of $O(q^2)$. In total, we get a decoding complexity of $O(q \log_2 q \cdot n \log_2 n)$. Also one can also consider a variant of the decoding procedure for the MAC setting. In (3), instead of taking a maximization over all elements of $\mathbb{F}_2^m$ elements, i.e., q elements, one can proceed sequentially. Namely, one finds for $k = 0,\dots,|S_J|-1$

$$\hat{u}_j[i_{k+1}] = \arg\max_{u \in \mathbb{F}_2}$$
$$\mathbb{P}\{U_j[i_{k+1}] = u | Y^n = y^n, U^{j-1} = \hat{u}^{j-1}, U_j[S_j^c] = \hat{u}_j[S_j^c],$$
$$U[i_1,\dots,i_k] = \hat{u}[i_1,\dots,i_k]\}.$$

This could only increase the probability of error by a factor $m = \log_2 q$, but this reduces the maximization search to at most m elements. However, one now needs to find the "conditional probabilities". If we compute them from the m-dimensional joint distribution, this may be equally costly. One may also work from the beginning with the conditional distributions, using a recursive procedure to compute them (leading to possible computational gains). This goes back to an onion-pealing approach (for the MAC or for the q-ary single-user channel), and although it may have desirable complexity attributes, one would have to evaluate and compare the entire performance (in particular error propagation and memory).

Also note that the problem of "shaping" arbitrary distributions, including discrete ones, can be approached using larger alphabet cardinalities (e.g., powers of two), and hence, the capacity of any memoryless channel can be achieved using polar coding schemes.

IV. Constellations for the AWGN channel

In this section, we consider the problem of finding discrete input distributions that serve as a good approximation of the Gaussian input distribution (with fixed variance), when the

[2] The mutual information of the MAC when the inputs are uniformly distributed

benchmark is the mutual information of the AWGN channel. Using previous sections, this leads to "polar coding schemes" for the AWGN channel.

A. Problem statement

From Section III, the input distribution of the polar coding schemes for single-user channels with alphabets $\mathbb{F}_q$, whether q is prime or a power of prime, is uniform. Motivated by the discussion of previous section, we consider input alphabets whose cardinality are powers of two.

Definition 1. For $m \in \mathbb{Z}_+$, we call a discrete random variable m-dyadic if its probability distribution p satisfies $p(x) = k2^{-m}$, for $k \in \mathbb{Z}_+$ and $x \in \mathrm{Supp}(p) \subset \mathbb{R}$.

Note that if X is m-dyadic, we have $|\mathrm{Supp}(p_X)| \leq 2^m$. Another way of defining m-dyadic random variables is to say that they are deterministic functions of m pure bits (i.e., i.i.d. Bernoulli half bits).

Consider now the following optimization problem. Let $v > 0$, $Z_v \sim \mathcal{N}(0, v)$, $m \in \mathbb{Z}_+$, and define

$$C_v(m) := \max_{X_m : \mathbb{E}X_m^2 \leq 1,\, X_m \text{ is } m\text{-dyadic}} I(X_m; X_m + Z_v). \quad (4)$$

Define also

$$C_v := 1/2 \log(1 + 1/v). \quad (5)$$

We would like to show that the gap $C_v - C_v(m)$ can be made small enough by choosing m small enough, so as to keep the complexity of the polar coding scheme defined in Section III-B manageable. Note that we could not find results about this problem in the literature.

B. The CLT approach

In [1], the following construction is proposed for the input,

$$G_m := \frac{1}{\sqrt{m}} \sum_{i=1}^{m} 2(X_i - 1/2).$$

Note that G_m is by construction m-dyadic and $\mathbb{E}G_m^2 \leq 1$. The support of G_m has cardinality $m + 1$, hence, the m-degrees of freedom are used in quantizing a "Gaussian" shape distribution. We now estimate how large the gap to capacity is for this input.

Lemma 1. *For any $v > 0$, there exists a constant $B > 0$ such that*

$$C_v - I(G_m; G_m + Z_v) \leq B/m$$

Proof: We have

$$C_v - I(G_m; G_m + Z_v) = D(G_m + Z_v || Z_{1+v})$$

and

$$G_m + Z_v = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} [2(X_i - 1/2) + \zeta_i] =: S_m,$$

where the ζ_i's are i.i.d. $\mathcal{N}(0, v)$. So the gap to capacity is exactly measured by the gap in divergence between a standardized sum of m random variables distributed as $2(X_1 - 1/2) + \zeta_1$ and the Gaussian distribution. In [8], this gap is precisely estimated, and it is shown that

$$D(S_m || Z_{1+v}) \leq B/m, \quad (6)$$

where B is a constant depending on the distribution of $2(X_1 - 1/2) + \zeta_1$, i.e., of a mixture of two Gaussian distributions (more precisely, it depends on the Poincaré constant of this distribution which is given by $v(1 + 2/v \exp(2/v))$). ■

The constant could possibly be improved, but the estimate seems accurate in the decay of $1/m$ (although for symmetric distributions a decay of $1/m^2$ could maybe be achieved). Hence, using the candidate G_m, and $q := 2^m$, we have a gap to capacity which decays like $1/\log_2(q)$. This is a slow decay in q. A small gap to capacity may then require q to be large (exponentially in m) and this may become critical for the complexity scaling discussed in Section III-B. We next show that this can be improved.

C. Exponential decay and quantile approach

The following improves the gap-to-capacity decay from $1/m$ to 2^{-m}.

Theorem 6. *For any v (with $\frac{1}{v}$ = SNR) and any $m \geq 1$*

$$C_v - C_v(m) \leq \frac{1}{v} 2^{-m}.$$

With such a decay, we can use Corollary 1 to cope with the complexity scaling of the MAC polar coding scheme, and approach the capacity of the AWGN channel with "low-complexity" codes, since a small gap to capacity requires $1/q = 2^{-m}$ to be small, i.e., m does not have to be very large. Indeed, we will see that m can be chosen to be fairly small for a reasonable gap (of course this also depends on the SNR). This Lemma is proved without an explicit construction of the input distribution; this is investigated in Section IV-C.

Proof of Theorem 6: Let $\xi_1, \ldots, \xi_{2^m}$ be i.i.d. $\mathcal{N}(0, 1)$, and let Q_m be uniformly distributed on $\{\xi_1, \ldots, \xi_{2^m}\} =: \xi$. Then, Q_m is m-dyadic (surely) and $\mathbb{E}Q_m^2 = 1$. Moreover,

$$\begin{aligned} & C_v - \mathbb{E}_\xi I(Q_m; Q_m + Z_v) = \mathbb{E}_\xi D(Q_m + Z_v || Z_{1+v}) \\ & \leq \mathbb{E}_\xi \chi^2(Q_m + Z_v || Z_{1+v}) \end{aligned} \quad (7)$$

where the last inequality uses the χ^2-distance, given by

$$\chi^2(Q_m + Z_v || Z_{1+v}) = \int_{\mathbb{R}} \frac{(p_m \star g_v(y) - g_{1+v}(y))^2}{g_{1+v}(y)} dy$$

where p_m denotes the random probability mass function of Q_m and g_v the Gaussian density with mean 0 and variance v. Note that $p_m \star g_v = \frac{1}{2^m} \sum_{i=1}^{2^m} g(y - \xi_i)$, and

$$\begin{aligned} & \mathbb{E}_\xi \int_{\mathbb{R}} \frac{(\mathbb{E}_{Q_m|\xi} g_v(y - Q_m) - g_{1+v}(y))^2}{g_{1+v}(x)} dy \\ & = 2^{-m} \int_{\mathbb{R}} \frac{\mathbb{E}_{Z_1} g_v(y - Z_1)^2 - g_{1+v}(y)^2}{g_{1+v}(x)} dy \\ & = 2^{-m} \left(\int_{\mathbb{R}} \frac{\mathbb{E}_{Z_1} g_v(y - Z_1)^2}{g_{1+v}(y)} dy - 1 \right) \end{aligned}$$

where $Z_1 \sim \mathcal{N}(0,1)$. Finally,

$$\int_{\mathbb{R}} \frac{\mathbb{E}_{Z_1} g_v(y-Z_1)^2}{g_{1+v}(y)} dy = \int_{\mathbb{R}} \frac{\int_{\mathbb{R}} g_v(y-z)^2 g_1(z)dz}{g_{1+v}(y)} dy \quad (8)$$

and since

$$g_v(y-z)^2 = \frac{1}{2\sqrt{\pi v}} g_{v/2}(y-z),$$

we have that $\int_{\mathbb{R}} g_v(y-z)^2 g_1(z)dz = \frac{1}{2\sqrt{\pi v}} g_{1+v/2}(y)$ and (8) is given by

$$\frac{1}{2\sqrt{\pi v}} \int_{\mathbb{R}} \frac{g_{1+v/2}(y)}{g_{1+v}(y)} dy = \frac{1+v}{v}$$

Therefore, $\mathbb{E}_{\xi}(C_v - I(Q_m; Q_m + Z_v)) \leq \frac{1}{v} 2^{-m}$, and by a standard probabilistic argument, there must exist a realization $\xi_1(\omega), \ldots, \xi_m(\omega)$, i.e., an input distribution uniformly distributed on these points, which satisfies the desired inequality. Indeed, the set of such realizations has high probability and one can show that there must exist realizations which also satisfy the power constraint. Also notice by tightening the bound in (7), one achieves a decay of $\log(1 + \frac{1}{v} 2^{-m})$. ■

One should stress that, although we use a probabilistic argument to show the existence of a good dyadic input distribution, as opposed to the problem of constructing good codes which may also use probabilistic arguments, the problem of finding a good dyadic distribution is much simpler (and can be investigated analytically or numerically). We provide here one such candidate.

Definition 2. [Gaussian quantile distribution] Let $D_m = \{k2^{-m} + 2^{-m-1} : k = 0, 1, \ldots, 2^m - 1\}$ and let U_m be drawn uniformly at random within D_n. We then define $Q_m^{(u)} = \Phi^{-1}(U_m)$, where Φ is the Normal c.d.f..

Analytical expansions suggests that $Q_m^{(u)}$ verifies the decay condition of Theorem 6. We show here numerical results in Figure 1, where the gap $C_v - I(Q_m^{(u)}; Q_m^{(u)} + Z_v)$ is plotted for different values of SNR $= 1/v$ and m.

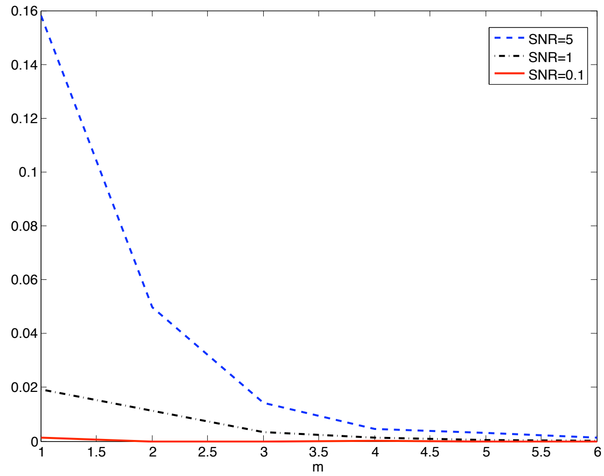

Fig. 1. Plots of $C_v - I(Q_m^{(u)}; Q_m^{(u)} + Z_v)$ for the Gaussian quantile distribution (cf. Definition 2) when m varies from 1 to 6 and for different values of SNR $= 1/v$.

V. Discussion

We have shown that the problem of finding good dyadic distributions as defined in Section IV-A can be significantly improved upon the CLT approach. The proposed constellation exploits the "spatial degree of freedom", i.e., it aims at carefully spacing uniform mass points to replicate the Gaussian mass distribution, rather than trying to approach a proper bell curve, as the CLT approach may do. On the other hand, in the spirit of the CLT approach, one may also look for a better way to discretize a Gaussian distribution using dyadic mass points. For example, the optimization problem which does not impose dyadic mass points, namely,

$$I_v(m) := \max_{X_m : \mathbb{E}X_m^2 \leq 1, |\text{Supp}(X_m)| \leq 2^m} I(X_m; X_m + Z_v)$$

has been studied [11], and a distribution for X_m is provided which ensures $C_v - I(X_m; X_m + Z_v) \leq 4(1 + 1/v)(1/(1 + v))^{2^m}$. This is a fast decay and one could quantize the obtained distribution to a dyadic one and estimate the new gap.

Regarding a more involved comparison of the different polar coding schemes discussed in this paper, the performance at a finite block length is of particular interest. A project is currently investigating this with numerical tests. It would then be interesting to compare the obtained results with other coding schemes such as LDPC or turbo codes, or the recently proposed scheme in [4].

References

[1] E. Abbe and E. Telatar, *Polar codes for the m-user MAC,* in Proc. 2010 International Zurich Seminar on Communications, Zurich, 2010. Available at arXiv:1002.0777v2.

[2] E. Arıkan, *Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,* IEEE Trans. Inform. Theory, vol. IT-55, pp. 3051–3073, July 2009.

[3] E. Arıkan and E. Telatar, *On the rate of channel polarization,* in Proc. 2009 IEEE Int. Symp. Inform. Theory, Seoul, pp. 1493–1495, 2009.

[4] A. R. Barron and A. Joseph, *Least squares superposition codes of moderate dictionary size, reliable at rates up to capacity*, in Proc. 2010 IEEE Int. Symp. Inform. Theory, Austin, pp. 275–279, 2010.

[5] A. Bennatan and D. Burshtein, *Design and analysis of nonbinary LDPC codes for arbitrary discrete-memoryless channels,* IEEE Trans. Inform. Theory, vol. 52, pp. 549–583, Feb. 2006.

[6] H. Cronie, *Coding and modulation for power and bandwidth efficient communication,* PhD thesis. University of Twente, Enschede, the Netherlands, September 2008.

[7] L. Duan, B. Rimoldi and R. L. Urbanke, *Approaching the AWGN channel capacity without active shaping,* in Proc. IEEE Int. Symp. on Information Theory, p. 374, June 1997.

[8] O. Johnson and A. Barron, *Fisher Information inequalities and the Central Limit Theorem* (2004), Probability Theory and Related Fields vol:129 , Issue:3 , Pages:391–409. DOI: 10.1007/s00440-004-0344-0.

[9] E. Şaşoğlu, E. Telatar, E. Arıkan, *Polarization for arbitrary discrete memoryless channels*, August 2009, arXiv:0908.0302v1.

[10] F.-W. Sun, H. C. A. van Tilborg, *Approaching capacity by equiprobable signaling on the Gaussian channel,* IEEE Trans. Inform. Theory, vol. 39, pp. 1714–1716, Sept. 1993.

[11] Y. Wu and S. Verdú, *The impact of constellation cardinality on Gaussian channel capacity*, Forty-Eighth Annual Allerton Conference on Communication, Control, and Computing, Monticello, September 2010.

[12] A. Viterbi, *Spread spectrum multiple access with binary PSK modulation can approach Shannon capacity for the aggregate Gaussian channel,* Qualcomm report, January 1986.