2019 IEEE Information Theory Workshop (ITW)

Channel Coding at Low Capacity

Mohammad Fereydounian
University of Pennsylvania
mferey@seas.upenn.edu

Mohammad Vahid Jamali
University of Michigan
mvjamali@umich.edu

Hamed Hassani
University of Pennsylvania
hassani@seas.upenn.edu

Hessam Mahdavifar
University of Michigan
hessam@umich.edu

***Abstract*—Low-capacity scenarios have become increasingly important in the technology of Internet of Things (IoT) and the next generation of mobile networks. Such scenarios require efficient and reliable transmission of information over channels with an extremely small capacity. Within these constraints, the performance of state-of-the-art coding techniques is far from optimal in terms of either rate or complexity. Moreover, the current non-asymptotic laws of optimal channel coding provide inaccurate predictions for coding in the low-capacity regime. In this paper, we provide the first comprehensive study of channel coding in the low-capacity regime. We will investigate the fundamental non-asymptotic limits for channel coding as well as challenges that must be overcome for efficient code design in low-capacity scenarios.**

I. Introduction

Low-capacity scenarios have become increasingly important in the technology of Internet of Things (IoT) and next generation of mobile networks. In particular, these scenarios have emerged in two extremes of wireless communications: narrowband and wideband communications. The former is widely considered for deploying IoT in cellular networks where massive number of users need to be served [1], and the latter models communication in the millimeter-Wave (mmWave) band which is one of the key innovations of the next generation of cellular networks (5G) [2]. From the channel modeling perspective, it turns out that users operating in these two different applications typically experience a very low signal-to-noise ratio (SNR). Therefore, studying fundamental limits as well as practical code construction is required to address the challenges of wireless system design for these emerging applications.

The Third Generation Partnership Project (3GPP) has introduced new features into the Long-Term Evolution (LTE) standard in order to integrate Internet of Things (IoT) into the cellular network. These new features, called Narrow-Band IoT (NB-IoT) and enhanced Machine-Type Communications (eMTC), have been introduced in the release 13 of LTE [1]. To ensure high coverage, the standard has to support coupling losses as large as 170 dB for these applications, which is approximately 20 dB higher than that of the legacy LTE. Tolerating such coupling losses requires reliable detection for a typical -13 dB of effective SNR, translated to capacity ≈ 0.03 bits/transmission. To enable reliable communication in such low-SNR regimes, LTE has adopted a legacy turbo code of rate $1/3$ as the mother code together with many repetitions. For NB-IoT, the standard allows up to 2048 repetitions to enable the maximum coverage requirements, thereby supporting effective code rates as low as 1.6×10^{-4} [1]. However, from a channel coding perspective, repeating a high-rate code to enable low-rate communication can be very sub-optimal.

Most of classical channel coding theory is centered on the design of point-to-point error-correcting codes, assuming an underlying channel with a certain capacity $C > 0$. However, since C is only asymptotically achievable, recently there has been a large body of work to study the *finite-length* performance: given a fixed block error probability p_e, what is the maximum achievable rate R in terms of the blocklength n? This question has been of interest to information theorists since the early years of information theory [3], and a precise characterization is provided in [4] as $R = C - \sqrt{\frac{V}{n}}Q^{-1}(p_e) + \mathcal{O}\left(\frac{\log n}{n}\right)$, where $Q(\cdot)$ is the tail probability of the standard normal distribution, and V is the channel dispersion. In recent years, this non-asymptotic law is specified up to the third and later to the fourth order for particular channels including BEC, BSC, and AWGN (see [5, Theorems 41,44], [6–10]). Such non-asymptotic laws have steered optimal code design for channels with moderate values of C. However, very little is known about optimal code design in the low-capacity regime, where channel capacity C could be extremely small (e.g. as small as $\mathcal{O}(1/n)$) and hence, the first and the second order terms of the law could be as small as the third order and the fourth order terms. In other words, for a given C, the prediction based on normal approximation, e.g., [4], after discarding the $\mathcal{O}(1)$ term, becomes more and more accurate as n grows large where the $\mathcal{O}(1)$ term becomes less and less relevant. However, when C is extremely small, even at moderate blocklengths, from coding application perspective, e.g., $C = 0.01$ and n being between 1000 and 10000, the $\mathcal{O}(1)$ term is still well comparable even with the first-order term nC leading to imprecise predictions by the normal approximation. More formally, such setting can be formulated as, for instance, $C = \mathcal{O}(1/n)$, which is particularly of interest from wideband communications perspective (see [11, Section 3]). Optimal code design in the low-capacity regime, defined explicitly in Section III, requires addressing various theoretical and practical challenges.

From the theoretical standpoint, channel variations in the low-capacity regime may be better approximated by different probabilistic laws rather than the ones used for typical channels. For instance, consider transmission over BEC(ϵ) with blocklength n. When the erasure probability ϵ is not very close to 1 (e.g., $\epsilon = 0.5$), the number of non-erased bits will be governed by the central limit theorem and behaves as $nC + \sqrt{n\epsilon(1-\epsilon)}Z$, where Z is the standard normal random variable. However, in the low-capacity regime, when the capacity $C = 1 - \epsilon$ is very small, although n is large, the number of channel non-erasures will not be large since a non-erasure occurs with small probability $1 - \epsilon$. In other words, the average number of non-erased bits is $n(1-\epsilon)$ which can be a constant or a number much smaller than n. Hence, the number of non-erasures will be best approximated by the law of rare events or the so-called Poisson convergence theorem rather than the central limit theorem.

From the design standpoint, we need to construct efficient codes with extremely low rates. Such constraints render the state-of-the-art codes and their advantages, in terms of decoding complexity and latency, inapplicable. For instance, it is well known that low-rate iterative codes have highly dense Tanner graphs which significantly deteriorates the performance (as there are many short cycles) as well as the computational complexity. Polar codes [12] can naturally be adapted to the low-rate regime. However, the current implementation of these codes suffers from relatively high computational complexity and latency, especially when such parameters are analyzed in terms of the number of information bits k rather than the blocklength n in the low-capacity regime.

This paper provides the first comprehensive study of channel coding at the low-capacity regime from both theoretical

978-1-5386-6900-6/19/$31.00 ©2019 IEEE

and code design viewpoints. We refer to [11] for a longer version of this paper with all the proofs. Section II provides the necessary background. Section III contains a precise definition of the low-capacity regime as well as fundamental non-asymptotic laws. Section IV considers various approaches to practical code design with simulation results and numerical comparisons.

II. Preliminaries

In this section, we will review the main concepts of channel coding in the non-asymptotic regime along with a brief review of previous works.[1] For an input alphabet $\mathcal{X}$ and an output alphabet $\mathcal{Y}$, a channel W can be defined as a conditional distribution on $\mathcal{Y}$ given $\mathcal{X}$. An (M, p_e)-code for the channel W is characterized by a message set $\mathcal{M} = \{1, 2, \cdots, M\}$, an encoding function $f_{enc} : \mathcal{M} \to \mathcal{X}$ and a decoding function $f_{dec} : \mathcal{Y} \to \mathcal{M}$ such that the *average* probability of error does not exceed p_e. Accordingly, an (M, p_e)-code for the channel W over n *independent channel uses* can be defined by replacing W with W^n in the definition. The blocklength of the code is defined as the number of channel uses and is denoted by n. For the channel W, the maximum code size achievable with a given error probability p_e and blocklength n is denoted by

$$M^*(n, p_e) = \max\left\{M \mid \exists (M, p_e)\text{-code for } W^n\right\}. \quad (1)$$

In this paper, we consider the binary erasure channel with erasure probability ϵ by BEC(ϵ) and binary symmetric channel with crossover probability δ by BSC(δ). The non-asymptotic expansion of $M^*(n, p_e)$ is given as [4,14]

$$\log_2 M^*(n, p_e) = nC - \sqrt{nV}Q^{-1}(p_e) + \mathcal{O}(\log_2 n), \quad (2)$$

where C is the channel capacity, V is the channel dispersion, and $Q(.)$ is the tail probability function of standard normal distribution. Note that the third order term in the non-asymptotic expansion of $M^*(n, p_e)$ depends on type of the channel (see [5, Theorems 41,44], [6], and [7–10]).

III. Fundamental Limits

The Low-Capacity Regime. Consider the transmission over a channel W with a capacity C. Let k denote the number of information bits to be sent and n denote the blocklength of the code. In *informal* terms, the low-capacity regime refers to scenarios in which the channel capacity C is very small. Indeed, To reliably communicate k bits, we clearly must have $n \geq k/C$ and thus n is a fixed and finite number but fairly large. For instance, consider the narrowband and wideband applications discussed in Section I, where the number of information bits k varies between few tens, in narrowband, to few thousands, in wideband (see [11, Section 3]), and the channel capacity C is typically below 0.05. This makes n to vary between few thousands to several tens of thousands. For instance, if $k = 50$ and $C = 0.02$, then the blocklength n is at least 2500.

In order to proceed with a *formal* definition of the low-capacity regime, we need to provide a formal characterization of the term "low" in the case where all the parameters such as C and n are assumed to be fixed an finite quantities (i.e. we consider the non-asymptotic setting).

Definition 1. *Let C and n be given fixed quantities. The transmission over a channel W with capacity C and blocklength n is called to be in the low-capacity regime if $C < n^{s-1}$, where $s \in [0, 1)$ is a tuning parameter quantifying what we consider as "low" in the application and its exact domain should be specified according to the channel.*

Remark 1. *Throughout the theorems and proofs, we define $\kappa := nC$ and consider n and C as "variables" which can take any given fixed value. This allows us to replace the condition $C < n^{s-1}$, for any $s \in [0, 2/3)$ (in the BSC case) with the condition $\kappa\sqrt{\kappa} = o(n)$ without abuse of the $o(.)$ notation (albeit n is originally a fixed rather than a growing value) which represents a comparison between the order of variables telling us which terms can be neglected. For instance, $\kappa = \mathcal{O}(1)$ (equivalently, $C = \mathcal{O}(1/n)$) or $\kappa = \mathcal{O}(\sqrt{n})$ (equivalently, $C = \mathcal{O}(1/\sqrt{n})$), etc.*

Why the laws should be different in the low-capacity regime? Let us now explain why the current non-asymptotic laws of channel coding provided in (2) are not applicable in the low-capacity regime. Consider transmission over BEC(ϵ) with blocklength n. When the erasure probability ϵ is not so large (e.g., $\epsilon = 0.5$), the number of channel non-erasures will be governed by the central limit theorem and behaves as $nC + \sqrt{n\epsilon(1-\epsilon)}Z$, where Z is the standard normal random variable. However, in the low-capacity regime, where the capacity $C = 1 - \epsilon$ is very small, the number of channel non-erasures will not be large, as the probability of non-erasure is very small. In other words, the expected number of non-erasures is $\kappa = n(1 - \epsilon)$ which is much smaller than n. In this case, the number of non-erasures is best approximated by the Poisson convergence theorem (i.e., the law of rare events) rather than the central limit theorem. Such behavioral differences in the channel variations will lead to totally different non-asymptotic laws, as we will see below.

Another reason for (2) being loose is that some of the terms that are considered as $\mathcal{O}(1)$ become significant in the low-capacity regime. E.g., we have $1/(\sqrt{n}C) = \sqrt{n}/(nC) = \sqrt{n}/\kappa$ which cannot be considered as $o(1)$ as κ is usually much smaller than n. As we will see, such terms can be captured by using sharper tail bounds.

Our approach. Note that extremely tight converse and achievability bounds for BEC and BSC have existed prior to [4,5] and stated as [5, Corollary 42, Theorem 43] for BEC and [5, Corollary 39, Theorem 40] for BSC. These bounds are in a raw implicit form. The novel contribution of [4,5] is using normal approximations and probability tail bounds to convert these implicit forms into explicit ones directly relating $\log_2 M^*(n, p_e)$ to n, p_e. This procedure works well for moderate values of C with respect to n but fails to provide accurate estimates in the low-capacity setting considered in this paper. In order to provide an accurate estimate, we need novel probabilistic laws which are, in some cases such as the BEC, totally different than what has been used before. Our approach can be summarized as follows: our starting points are the same as [4,5], i.e., we start with [5, Corollary 42, Theorem 43] for BEC and [5, Corollary 39, Theorem 40] for BSC, but our analysis is based on Poisson approximations (for BEC) and much tighter probability tail bounds (for BSC) which are specifically perfect for the low-capacity regime but not necessary for moderate values of C. These novel approaches in analysis lead to the low-capacity coding bounds for BEC and BSC stated in the following subsections.

A. The Binary Erasure Channel

The following theorem provides lower and upper bounds for the best achievable rate (see (1)) in terms of n, p_e, ϵ, and $\kappa := n(1 - \epsilon)$. For the Poisson distribution, we use $\mathcal{P}_\lambda(x)$ to denote

$$\mathcal{P}_\lambda(x) = \Pr\left\{X < x\right\}, \qquad \text{where } X \sim \text{Poisson}(\lambda). \quad (3)$$

Theorem 1 (Non-Asymptotic Coding Bounds for the Low–Capacity BEC)**.** *Consider transmission over* BEC(ϵ) *in the low-capacity regime and let $\kappa = n(1 - \epsilon)$. Then,*

$$M_1 \leq M^*(n, p_e) \leq M_2,$$

where M_1 is the solution of

$$\mathfrak{P}_1(M_1) + \alpha\sqrt{\mathfrak{P}_1(M_1)} - p_e = 0, \quad (4)$$

[1] For more details, we refer the reader to [13] for an excellent review.

and M_2 is the solution of

$$\mathfrak{P}_2(M_2) - \alpha\sqrt{\mathfrak{P}_2(M_2)} - \alpha\sqrt{\mathcal{P}_\kappa(\log_2 M_2)} - p_e = 0, \quad (5)$$

and

$$\mathfrak{P}_1(M_1) = \mathcal{P}_\kappa(\log_2 M_1) + M_1 e^{-\kappa/2}\left(1 - \mathcal{P}_{\kappa/2}(\log_2 M_1)\right),$$
$$\mathfrak{P}_2(M_2) = \mathcal{P}_\kappa(\log_2 M_2) - \frac{e^\kappa}{M_2}\mathcal{P}_{2\kappa}(\log_2 M_2),$$
$$\alpha = \frac{\sqrt{2}}{\epsilon^{3/2}}\left(1 + 2\sqrt{\frac{3}{\epsilon\kappa}}\right)(\sqrt{e} - 1)(1 - \epsilon).$$

The bounds in Theorem 1 are tight and can be simply computed numerically (see [11, Appendix]). The bounds are expressed merely in terms of $\kappa := n(1-\epsilon)$ rather than n.

B. The Binary Symmetric Channel

Unlike BEC, the non-asymptotic behavior of coding over BSC can be well approximated in the low-capacity regime by the central limit theorem (e.g., Berry-Essen theorem). In the following we briefly explain the reason. Consider transmission over BSC(δ) where the value of δ is close to $1/2$. The capacity of this channel is $1 - h_2(\delta)$, where $h_2(x) := -x\log_2(x) - (1-x)\log_2(1-x)$, and we denote $\kappa = n(1 - h_2(\delta))$. Note that when $\delta \to 1/2$ one can write $\delta \approx 1/2 - \sqrt{\kappa/n}$ by using the Taylor expansion of the function $h_2(x)$ around $x = 1/2$. Transmission over BSC(δ) can be equivalently modeled as follows: (i) With probability 2δ we let the output of the channel be chosen according to Bernoulli($1/2$), i.e., the output is completely random and independent of the input, and (ii) with probability $1-2\delta$ we let the output be exactly equal to the input. In other words, the output is completely noisy with probability 2δ (call it the noisy event) and completely noiseless with probability $1 - 2\delta$ (call it the noiseless event). As $\delta \to 1/2$, then the noiseless event is a *rare event*. Now, assuming n transmissions over the channel, the expected number of noiseless events is $n(1-2\delta) \approx \sqrt{n\kappa}$. Similar to BEC, the number of rare noiseless events follows a Poisson distribution with mean $n(1-2\delta)$ due to the Poisson convergence theorem. However, as the value of $n(1-2\delta) \approx \sqrt{n\kappa}$ is large, the resulting Poisson distribution can also be well approximated by the Gaussian distribution due to the central limit theorem (note that Poisson(m) can be written as the sum of m independent Poisson(1) random variables).

As mentioned earlier, central limit laws are the basis for deriving the laws of the form (2) which are applied to the settings where the capacity is not small. However, for the low-capacity regime, considerable extra effort is required in terms of sharper arguments and tail bounds to work out the constants correctly.

Theorem 2 (Non-Asymptotic Coding Bounds for the Low–Capacity BSC)**.** *Consider transmission over BSC(δ) in the low-capacity regime and let $\kappa = n(1 - h_2(\delta))$. Then,*

$$\log_2 M^*(n, p_e) = \kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}}\cdot\sqrt{\kappa}\,Q^{-1}(p_e) + \frac{1}{2}\log_2\kappa - \log_2 p_e + \mathcal{O}(\log\log\kappa). \quad (6)$$

The superiority of this novel derivation compared to the state-of-the-art BSC bound [5, Theorem 41] in the low-capacity regime is numerically shown in Section IV-C.

IV. Practical Code Designs and Simulation Results

As we need to design codes with extremely low rates, some of the state-of-the-art codes may not be directly applicable. A notable instance is the class of iterative codes, e.g., Turbo or LDPC codes. It is well known that decreasing the design rate of iterative codes results in denser decoding graphs which further leads to highly complex iterative decoders with poor performance. In order to circumvent this issue, the current practical designs, such as the NB-IoT code design, use repetition coding, i.e., a low rate repetition code is concatenated with a powerful moderate-rate code. In Section IV-A, we will provide fundamental trade-offs between the number of repetitions and performance of the code, and show that, even though repetition leads to efficient implementations, the rate loss through many repetitions will result in codes with mediocre performance. In Section IV-B, we will study the behavior of polar coding on low-capacity channels. As we will see, polar coding is advantageous in terms of distance, performance and implicit repetition, however, it can be further simplified for practical applications. Throughout this section, we will consider code design for the class of binary memoryless symmetric (BMS) channels.

A. Repetition Coding

As mentioned above, repetition is a simple way to design practical low-rate codes that exploit the power of state-of-the-art designs. Let r be a divisor of n, where n denotes the length of the code. Repetition coding consists in designing first a smaller outer code of length n/r and repeat each of its code bits r times (i.e., the inner code is repetition). The length of the final code is $n/r \cdot r = n$. This is equivalent to transmitting the outer code over the *r-repetition channel*, W^r, which takes a bit as input, and outputs an r-tuple that is the result of passing r copies of the input bit independently through the original channel W, e.g., if W is BEC(ϵ) then its corresponding r-repetition channel is $W^r = \text{BEC}(\epsilon^r)$. The main advantage of repetition coding is the reduction in computational complexity and latency (especially if r is large). This is because the encoding/decoding complexity is effectively reduced to that of the outer code.

The outer code has to be designed for reliable communication over the channel W^r. If r is sufficiently large, then the capacity of W^r will not be low anymore. In this case, the outer code can be picked from off-the-shelf practical codes designed for channels with moderate capacity values (e.g., iterative or polar codes). While this looks promising, one should note that the main drawback of repetition coding is the loss in capacity. In general, we have $C(W^r) \le rC(W)$ and the ratio vanishes by growing r. As a result, if r is very large then repetition coding might suffer from an unacceptable rate loss. Thus, the main question that we need to answer is: how large r can be made such that the rate loss is still negligible?

We note that the overall capacity corresponding to n channel transmissions is $nC(W)$. With repetition coding, the capacity will be reduced to $n/r \cdot C(W^r)$ since we transmit n/r times over the channel W^r. For any $\beta \in [0, 1]$, we ask what is the largest repetition size r_β such that

$$\frac{n}{r_\beta}C(W^{r_\beta}) \ge \beta n C(W). \quad (7)$$

Theorem 3 (Maximum Repetition Length for BEC)**.** *If $W =$ BEC(ϵ), then for the largest repetition size r_β that satisfies* (7)*, we have*

$$\frac{n(1-\epsilon)\ell}{2\left(1 - \frac{\beta}{\ell}\right)}\cdot\left(\frac{\beta}{\ell}\right)^2 \le \frac{n}{r_\beta} \le \frac{n(1-\epsilon)\ell}{2\left(1 - \frac{\beta}{\ell}\right)}, \quad (8)$$

where $\ell = -\frac{\ln\epsilon}{1-\epsilon}$. Equivalently, assuming $\kappa = n(1-\epsilon)$, (8) *becomes*

$$\frac{\kappa}{2(1-\beta)}\cdot\beta^2(1 + \mathcal{O}(1-\epsilon)) \le \frac{n}{r_\beta} \le \frac{\kappa}{2(1-\beta)}(1 + \mathcal{O}(1-\epsilon)).$$

n	1024	4096	16364
$d_{\min}(d_{\min}/n)$	128 (1/8)	512 (1/8)	2048 (1/8)

TABLE I: Minimum distance of a polar code constructed for $k = 40$ over various channels with capacity 0.02.

Remark 2. *Going back to the results of Theorem 1, in order to obtain similar non-asymptotic guarantees with repetition-coding, a necessary condition is that the total rate loss due to repetition is $\mathcal{O}(1)$, i.e.,*

$$\frac{n}{r_\beta}C(W^{r_\beta}) = nC(W) + \mathcal{O}(1).$$

If $W = BEC(\epsilon)$ and $\kappa = n(1-\epsilon)$, then the necessary condition implies plugging $\beta = 1 - \mathcal{O}(1/\kappa)$ into (7). *Moreover, from Theorem 3 we can conclude that, when ϵ is close to* 1*, the maximum allowed repetition size is $\mathcal{O}\left(n/\kappa^2\right)$. Equivalently, the size of the outer code can be chosen as $\mathcal{O}(\kappa^2)$.*

We conclude that, as having a negligible rate loss implies the repetition size to be at most $\mathcal{O}(n/\kappa^2)$, the outer code has to be designed for a BEC with erasure probability at least $\epsilon^{\mathcal{O}(n/\kappa^2)} = 1 - \mathcal{O}(1/\kappa)$. This means that the outer code should still have a low rate even if κ is as small as few tens. Thus, the idea of using e.g., iterative codes as the outer code and repetition codes as the inner code will lead to an efficient low-rate design only if we are willing to tolerate non-negligible rate losses. We refer to Section IV-C for a numerical case study on repetition coding.

It turns out that the binary erasure channel has the smallest rate loss due to repetition among all the BMS channels. This property has been used in the following theorem to provide an upper bound on r_β for any BMS channel.

Theorem 4 (Upper Bound on Repetition Length for any BMS)**.** *Among all BMS channels with the same capacity, BEC has the largest repetition length r_β that satisfies* (7). *Hence, for any BMS channel with capacity C and $\kappa = nC$, we have*

$$\frac{n}{r_\beta} \geq \frac{\kappa}{2(1-\beta)}\beta^2(1 + \mathcal{O}(1-C)).$$

Remark 3. *Similar to Remark 2, we conclude that for any low-capacity BMS channel, in order to have the total rate loss of order $\mathcal{O}(1)$, the repetition size should be at most $\mathcal{O}(n/\kappa^2)$.*

B. Polar Coding at Low Capacity

We show in this section that polar construction provides several coding advantages, in terms of both performance and complexity, in the low-capacity regime. We also show that, to make polar codes a suitable candidate for practice, we need to carefully adapt their encoding and decoding operations.

The generator matrix of polar codes of length $n = 2^m$ is based on choosing k rows of the matrix $G_n = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes m}$, corresponding to the best k "synthetic" channels [12].

High Minimum Distance at Low-Capacity. If the channel W has low capacity, then clearly any good (i.e., noiseless) synthetic channel requires a lot of plus operations. As a result, for all the k best synthetic channels the Hamming weight of the corresponding row in G_n is very high. Hence, the resulting polar code will have a high minimum distance. Table I provides the minimum distance of the polar code for various channels and lengths. The channels are BAWGN, BEC, BSC all with capacity 0.02. We have constructed polar codes for these channels with $k = 40$. Indeed, polar and Reed-Muller codes become very similar in the low-capacity regime.

Polar Coding Does Optimal and Implicit Repetition at Low-Capacity. We have shown in Section IV-A that the maximum allowed repetition size to have a negligible capacity loss is $\mathcal{O}(n/\kappa^2)$. We will show in this section that at low-capacity, the polar construction is enforced to have $\mathcal{O}(n/\kappa^2)$ repetitions. In other words, the resulting polar code is equivalent to a smaller polar code of size $\mathcal{O}(\kappa^2)$ followed by repetitions. Consequently, the encoder and decoder of the polar code could be implemented with much lower complexity taking into account the implicit repetitions. That is, the encoding can be reduced to $n + \mathcal{O}(\kappa^2 \log \kappa)$ and the decoding complexity using the list successive cancellation (SC) decoder with list size L is reduced to $n + \mathcal{O}(L\kappa^2 \log \kappa)$. Recall that the original implementation of polar codes requires $n \log n$ encoding complexity and $\mathcal{O}(Ln \log n)$ decoding complexity. Moreover, as the repetition steps can all be done in parallel, the computational *latency* of the encoding and decoding operations can be reduced to $\mathcal{O}(\kappa^2 \log \kappa)$ and $\mathcal{O}(L\kappa^2 \log \kappa)$, respectively. To further reduce the complexity, simplified SC decoders can be invoked, and are crucial for making polar codes a suitable candidate for practice.

Theorem 5. *Consider using a polar code of length $n = 2^m$ for transmission over a BMS channel W. Let $m_0 = \log_2(4\kappa^2)$ where $\kappa = nC(W)$. Then any synthetic channel $W_n^{(i)}$ whose Bhattacharyya value is less than $1/2$ has at least m_0 plus operations in the beginning. As a result, the polar code constructed for W is equivalent to the concatenation of a polar code of length (at most) 2^{m_0} followed by 2^{m-m_0} repetitions.*

Remark 4. *Note that from Theorem 5, polar codes automatically perform repetition coding with $\mathcal{O}(n/\kappa^2)$ repetitions, where $\kappa = nC$. This matches the necessary (optimal) number of repetitions given in Remark 2 and 3.*

C. Simulation Results

For the BEC, Figure **1** compares the lower and upper bounds obtained from Theorem 1 with the predictions of Formula (2). We have also plotted the performance of random linear codes as well as polar codes (with successive list decoding, $L = 16$ and 6 cyclic redundancy check (CRC) bits [15]). The setting considered in Figure **1** is as follows: We intend to send $k = 40$ information bits over the BEC(ϵ). The desired error probability is $p_e = 10^{-2}$. For erasure values between 0.96 and 1, Figure **1** plots bounds on the smallest (optimal) blocklength n needed for this scenario as well as the smallest length required by polar codes. Note that in order to compute a lower bound on the shortest length from Theorem 1, we should fix $M^*(n, p_e)$ to $k = 40$ and search for the smallest n satisfying (5) with $\kappa = n(1-\epsilon)$ and $p_e = 0.01$.

As we see in Figure **1**, the lower and upper bounds predicted from Theorem 1 are very close to each other. The performance of random linear codes is very close to the upper bound which is natural as the upper bound uses a random coding achievability argument. As expected, the prediction obtained from Formula (2) is not precise in the low-capacity regime and it becomes worse as the capacity approaches zero.

Figure **1** also includes bounds and predictions for the BSC under the same setting (i.e. $k = 40$ and $p_e = 0.01$). We have compared in Figure **1**, the predictions from Theorem 2 and Formula (2) together with the upper and lower bounds which are directly computed from Random coding union bound (for instance, see [5, Corollary 39]) and a general tight converse bound for BSC (see e.g. [5, Theorem 40]), respectively. Note that the true value of n lies between these two bounds. Theorem 2 and Formula (2) are both estimating that true value based on these upper and lower bounds. In this way, Figure **1** shows that, as we expected, the prediction from Formula (2) (obtained in [4]) is quite imprecise in the low-capacity regime particularly in comparison to the prediction from Theorem 2.

Figure **2** compares the performance of polar codes with repeated LTE Turbo codes over the binary-input additive

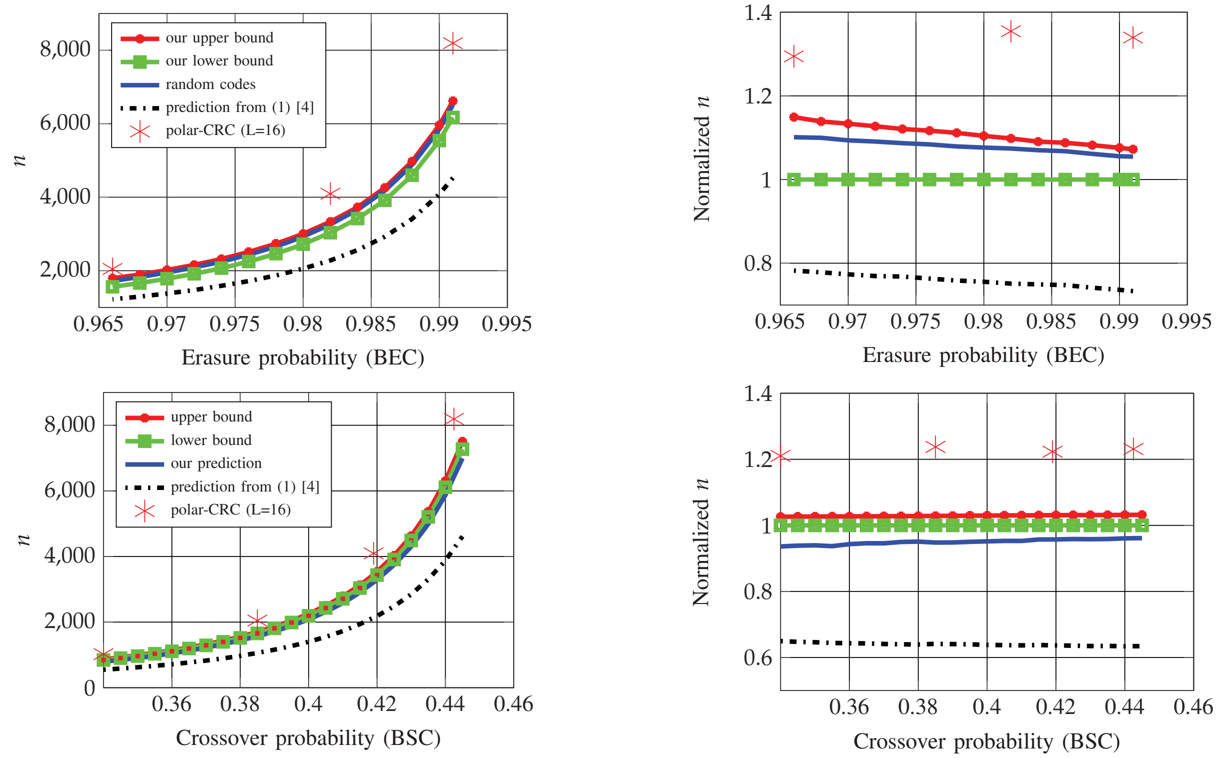

Figure 1: Comparison for the low-capacity BEC (top plots) and BSC (bottom plots). The number of information bits is $k = 40$ and the target error probability is $p_e = 10^{-2}$. Each plot on the right is a "normalized" version of its left counterpart, i.e., all the blocklengths n in the left plots are normalized by the value of the lower bound and are plotted in the right. Also, for each of the plots on the right we have used the same legend entries as their left counterpart.

white Gaussian noise channel. Here, we intend to send $k = 40$ information bits. The polar-CRC code has length 8192, and the Turbo-repetition scheme has the $(120, 40)$ mother code of rate $1/3$ as the outer code which is repeated 68 times (the total length is $68 \times 120 = 8160$). In the considered $(8192, 40)$ polar code, a repetition factor of 4 is implicitly enforced by the construction, as predicted by Theorem 5. Hence, the polar coding scheme is actually a $(2048, 40)$ polar code with 4 repetitions. We note from Section I that repetition of the LTE code for data channel, in this case the Turbo code of rate $1/3$, is the proposed code design in the NB-IoT standard. For these two choices of code designs, the block error probability is plotted with respect to E_b/N_0 in Figure **2**. As we see from the figure, the waterfall region of Turbo-repetition is almost 4 dB away from that of the polar code. This is mainly due to the many repetitions that must be invoked in the repeated Turbo code to provide the low rate design. Consequently, this results in capacity loss and significantly degraded performance for Turbo-repetition scheme comparing to the carefully designed polar code.

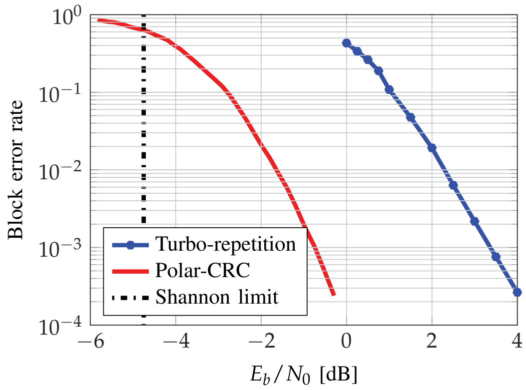

Figure 2: Comparison for the low-capacity BAWGN between polar codes and Turbo-repetition codes. The Shannon limit for this setting is -4.75 dB.

REFERENCES

[1] R. Ratasuk, N. Mangalvedhe, Y. Zhang, M. Robert, and J.-P. Koskinen, "Overview of narrowband IoT in LTE Rel-13," in *Proc. IEEE Conf. Standard Commun. Netw.*, Berlin, Germany, Oct./Nov. 2016, pp. 1–7.

[2] "5G; Study on scenarios and requirements for next generation access technologies," 3GPP TR 38.913 version 14.3.0 Release 14, 2017.

[3] V. Strassen, "Asymptotische abschätzungen in Shannon's informationstheorie," in *Trans. 3rd Prague Conf. Inf. Theory*, pp. 689–723, 1962.

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

[5] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, Prinston University, 11 2010.

[6] V. Y. F. Tan and M. Tomamichel, "The third-order term in the normal approximation for the awgn channel," *IEEE Trans. Info. Theory*, vol. 61, no. 5, pp. 2430–2438, May 2015.

[7] T. Erseghe, "Coding in the finite-blocklength regime: Bounds based on laplace integrals and their asymptotic approximations," *IEEE Trans. Info. Theory*, vol. 62, no. 12, pp. 6854–6883, Dec 2016.

[8] P. Moulin, "The log-volume of optimal codes for memoryless channels, asymptotically within a few nats," *IEEE Trans. Info. Theory*, vol. 63, no. 4, pp. 2278–2313, April 2017.

[9] J. Scarlett, A. Martinez, and A. G. i. Fabregas, "Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations," *IEEE Trans. Info. Theory*, vol. 60, no. 5, pp. 2647–2666, 2014.

[10] W. Yang, A. Collins, G. Durisi, Y. Polyanskiy, and H. V. Poor, "Beta–Beta bounds: Finite-blocklength analog of the golden formula," *IEEE Trans. Info. Theory*, vol. 64, pp. 6236–6256, 2018.

[11] M. Fereydounian, M. V. Jamali, H. Hassani, and H. Mahdavifar, "Channel coding at low capacity," *arXiv preprint arXiv:1811.04322*, 2018.

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

[13] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *arXiv:1504.02608v1 [cs.IT]*, 2015.

[14] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Info. Theory*, vol. 55, no. 11, pp. 4947–4966, 2009.

[15] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Info. Theory*, vol. 61, no. 5, pp. 2213–2226, 2015.