Channel Coding at Low Capacity

Mohammad Fereydounian, *Graduate Student Member, IEEE,* Hamed Hassani, *Member, IEEE,*
Mohammad Vahid Jamali, *Member, IEEE,* and Hessam Mahdavifar, *Member, IEEE*

Abstract—Low-capacity scenarios have become increasingly important in the technology of the Internet of Things (IoT) and the next generation of wireless networks. Such scenarios require efficient and reliable transmission over channels with an extremely small capacity. Within these constraints, the state-of-the-art coding techniques may not be directly applicable. Moreover, the prior work on the finite-length analysis of optimal channel coding provides inaccurate predictions of the limits in the low-capacity regime. In this paper, we study channel coding at low capacity from two perspectives: fundamental limits at finite length and code constructions. We first specify what a low-capacity regime means. We then characterize finite-length fundamental limits of channel coding in the low-capacity regime for various types of channels, including binary erasure channels (BECs), binary symmetric channels (BSCs), and additive white Gaussian noise (AWGN) channels. From the code construction perspective, we characterize the optimal number of repetitions for transmission over binary memoryless symmetric (BMS) channels, in terms of the code blocklength and the underlying channel capacity, such that the capacity loss due to the repetition is negligible. Furthermore, it is shown that capacity-achieving polar codes naturally adopt the aforementioned optimal number of repetitions.

Index Terms—Internet of things, polar codes, finite-length coding bounds, achievability and converse bounds, low rate coding, low SNR, Poisson law.

I. Introduction

Error-correcting codes are often designed assuming an underlying channel with a certain capacity $C > 0$. In order to understand how optimal the designed codes are, studying the finite-length fundamental limits becomes relevant, i.e., given a fixed block error probability p_e, what is the maximum achievable rate R in terms of the blocklength n? There has been a large body of work in the past decade to study the fundamental limits of finite-length channel coding relating p_e, R, and n together. This has been of interest to information theorists since the early years of information theory

[3], [4], and a precise characterization is provided in [5] as $R = C - \sqrt{V/n}\, Q^{-1}(p_e) + \mathcal{O}\left(\log n/n\right)$, where C is the channel capacity, $Q(\cdot)$ is the tail probability of the standard normal distribution, and V is a characteristic of the channel referred to as channel dispersion. In recent years, this finite-length analysis is further enhanced to include up to the third and later to the fourth order for particular channels including BEC, BSC, and AWGN (see [6, Theorems 41,44], [7]–[11]).

In general, the fundamental question of what is achievable in the finite-length regime has been answered for various types of channels and up to several orders of approximation in the *moderate-capacity* regime, where the higher order terms of approximating R are significantly smaller than the first few terms. In this paper, we consider cases where the capacity C is extremely small where the first-order (i.e., C) and/or the second-order terms are as small as the higher-order terms. In general, as we will see throughout this paper, designing optimal channel codes in such a *low-capacity* regime (which we explicitly specify in Section III) and understanding how far they are from what is fundamentally achievable require addressing various theoretical and practical challenges.

From the code construction perspective, some of the state-of-the-art codes may not be directly applicable in the extremely low-rate regime. A notable instance is the class of iterative codes, e.g., turbo [12] or low-density parity-check (LDPC) codes [13], [14]. 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. To circumvent this issue, the current practical designs use repetition coding. In particular, a low-rate repetition code is concatenated with a powerful moderate-rate code. Although repetition leads to efficient implementations, the rate loss through many repetitions may become significant. This implies that a comprehensive analysis is necessary to understand the optimality of coded repetition schemes in the low-capacity regime.

A. Problem Motivation

Low-capacity scenarios have become increasingly important in the technology of the Internet of Things (IoT) and the next generation of wireless networks. The Third Generation Partnership Project (3GPP) has introduced new features into the standard in order to integrate IoT into the cellular network. These new features, called Narrow-Band IoT (NB-IoT) and enhanced Machine-Type Communications (eMTC), were introduced in the release 13 of 3GPP standard and have been evolving since then. The aim of these features is to enable deploying IoT in cellular networks where a massive number of users need to be served [15]. From the channel

modeling perspective, it turns out that users operating in these modes typically experience very low signal-to-noise ratios (SNRs). In particular, 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 standard. Tolerating such coupling losses requires reliable detection for a typical -13 dB of effective SNR [15], [16], translated to capacity ≈ 0.03 bits/transmission. To enable reliable communications in such low-SNR regimes, the standard has adopted a legacy turbo code of moderate rate, i.e., rate $1/3$, in eMTC and NB-IoT (uplink) as the mother code together with many repetitions. 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} [15]. However, as mentioned earlier, such repetition schemes may result in a significant rate loss. In general, studying finite-length fundamental limits as well as designing practical code constructions are necessitated to address the challenges of wireless system design for such emerging applications.

Communication in low-capacity regimes is also relevant in deep-space communication. In addition to the limited capacity, deep-space communication also suffers from catastrophic link loss and severe signal attenuation. Hence, sophisticated code concatenation designs are often required in order to combat these design challenges. An overview of code designs adopted for various historical deep-space missions can be found in [17]. Designing efficient coding techniques to enhance the performance of deep-space communication is still an ongoing and open area of research that necessitates further attention given the importance of the targeted applications [18].

B. Related Work

Following the earlier work of Polyanskiy et al. [5], fundamental limits at finite length were later studied for various other types of channels beyond BECs, BSCs, and AWGN channels, including block-fading channels [19], [20], multiple-antenna channels [21], and multiple access channels [22], [23]. This has also motivated studying finite-length analysis in other related settings including lossy compression [24], Slepian-Wolf coding [23], [25], covert communications [26], [27], and coding with side-information [28], among others.

Another line of work in the literature is concerned with the application of saddlepoint approximations to efficiently compute rather complicated expressions such as random-coding union bound [10], [29]–[33]. To this end, [29] derived saddlepoint approximations of random-coding bounds to the decoding error probability with maximum-metric mismatched decoders allowing for accurate and simple numerical evaluations. In [30], a single-letter saddlepoint approximation, that is shown to be asymptotically tight for both fixed and varying rates, is presented for random-coding union bound of Polyanskiy *et al.* [5] for i.i.d. random coding over discrete memoryless channels. Moreover, saddlepoint approximations of the meta-converse (hypothesis-testing) lower bound and random-coding union upper bound of channel coding minimum error probability are derived in [31] for symmetric memoryless channels in a wide range of system parameters.

In a related line of work, *very noisy channels* (VNCs) are defined and studied. The notion of VNCs was first defined by Reiffen in [34] by specifying certain conditions on the channel transition probability. Later, Gallager computed exponent-rate functions for random coding and convolutional codes in [35], and Majani [36] carried out a comprehensive study of VNCs. VNCs are also relevant in Poisson photon channels, modeling direct detection optical communication channels when they are approximated by binary-input binary-output discrete memoryless channels [37]. Recently, Sakai *et al.* [38] derived finite-length laws for channel coding over continuous-time Poisson channels. Also, very recently, Wagner *et al.* established that feedback neither improves the second-order coding rate for very noisy discrete memoryless channels [39] nor their high-rate error exponent or moderate deviations performance [40]. Although the low-capacity setting in this paper shares similar motivations to that of VNCs, the characterization of the low-capacity regime for a channel is fundamentally different. To clarify this difference, note that there is no notion of block-length in the formulation of VNCs. However, our definition of low-capacity channels directly relates the low-capacity regime to the blocklength. More precisely, according to what will be discussed in Section III-A, a channel with a fixed capacity C may not be at low capacity for a given blocklength, but may fall in the low-capacity regime for a shorter blocklength. Therefore, a VNC may or may not be a low-capacity channel necessarily.

For code construction, the focus of this paper is on the class of binary memoryless symmetric (BMS) channels. Asymptotically, state-of-the-art polar codes, introduced by Arıkan [41], are the first class of provably capacity-achieving codes with explicit constructions as well as low-complexity encoding and decoding. Furthermore, their construction method is rate-adaptive, allowing constructing codes of rate k/n for $k = 0, 1, 2 \ldots, n$, where n is the block length. While this makes them a natural choice for low-capacity regimes, they have not been particularly studied in very low-rate regimes when the number of information bits k is much smaller than n.

C. Our Contributions

In this paper, we provide a specific formulation of low-capacity regimes from a finite-length analysis perspective. We then provide fundamental non-asymptotic laws of channel coding in the low-capacity regime for a diverse set of channels with practical significance, namely, BEC, BSC, and AWGN channels. We observe that channel variations in the low-capacity regime can be better approximated by different probabilistic laws rather than the ones used for channels with moderate capacities. In particular, for BEC channels, we show that the behavior of channel variations in the low-capacity regime can be better approximated through the Poisson convergence theorem that studies the occurrence of rare events. This is basically intuitive noticing that "non-erasure" in a BEC with a very small capacity is a rare event. This phenomenon in the low-capacity BEC changes the relative significance of order terms in the classical expansions of the best achievable rate [5] in such a way that the higher order

terms are then comparable to the lower ones and hence leading to an imprecise approximation. Incorporating Poisson laws in this paper rather than the Gaussian laws used in the classical expansions, however, results in a more accurate approximation. The inaccuracy of the classical expansion in the low-capacity BSC is as well due to the aforementioned change in the relative significance of the order terms but unlike BEC, the Poisson laws do not govern the behavior of the BSC in the low-capacity regime and are not applicable here. As our analysis shows, employing the Gaussian laws with a sharper analysis and more precise tail bounds (e.g., the bound obtained by Talagrand [42]) can circumvent this issue and lead to a different and more accurate expansion. For an AWGN channel, it turns out that the low-capacity expansion can be seen as a term-by-term limit of the existing expansion in the moderate-capacity regime. Proving this observation in Section III-D is our contribution to the AWGN channel case. This leaves no necessity for a numerical evaluation of the AWGN case.

From the code construction perspective, assuming transmission over a BMS channel, repetition is often considered as a straightforward method to design practical low-rate binary codes that utilize the power of state-of-the-art binary code designs at a moderate rate while keeping the complexity low. This is mainly due to the fact that the encoding/decoding complexity of a concatenation scheme with inner repetition is effectively reduced to that of the outer code with a significantly shorter length. Thus, a major question is how large the number of repetitions can be such that the capacity loss due to the repetition is negligible? To answer this fundamental question, we characterize the optimal number of repetitions, in terms of the code blocklength and the underlying channel capacity. As mentioned earlier, polar codes are very appealing for low-capacity regimes due to their rate-adaptive structure. In this regard, we prove that the polarization transform implicitly induces the aforementioned optimal number of repetitions that we characterize in the low-capacity regime. This means the resulting low-rate polar codes naturally adopt the optimal number of repetitions.

Our approximations of the fundamental limits in the low-capacity regime for the BEC and BSC cases are numerically evaluated and compared with the most well-known moderate-capacity estimation [5] and an all-rate estimation known as the saddlepoint approximation [30].

D. Content Organization

The rest of this paper is organized as follows. In Section II, we provide the necessary background. In Section III, we formally define the low-capacity regime and provide non-asymptotic laws of channel coding. Section IV is devoted to the code design. The numerical results are discussed in Section V. Section VI presents the conclusion and some future directions. Finally, for all proofs of the theorems and their intermediate lemmas, we refer to the arXiv version of this paper [2].

II. PRELIMINARIES

A. Finite-Length Analysis

In this section, we will review the main concepts of channel coding in the finite-length regime, sometimes referred to as the non-asymptotic regime in the literature, 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 $f_{enc} : \mathcal{M} \rightarrow \mathcal{X}$, and a decoding function $f_{dec} : \mathcal{Y} \rightarrow \mathcal{M}$ such that the *average* probability of error does not exceed p_e, that is[2]

$$\frac{1}{M} \sum_{m \in \mathcal{M}} W\left(\mathcal{Y} \setminus f_{dec}^{-1}(m) \,\middle|\, f_{enc}(m)\right) \leq p_e. \quad (1)$$

We consider p_e to be a fixed given constant in $(0, 1)$. 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 similarly 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

$$M^*(n, p_e) = \max\left\{M \mid \exists (M, p_e)\text{-code for } W^n\right\}. \quad (2)$$

In this paper, we consider three classes of channels that vary in nature:

- BEC(ϵ): binary erasure channel with erasure probability $\epsilon \in (0, 1)$.
- BSC(δ): binary symmetric channel with crossover probability $\delta \in (0, 1)$.
- AWGN(η): additive white Gaussian noise channel with signal-to-noise ratio $\eta \in (0, \infty)$.

Next, we mention the well-known finite-length expansion for a discrete memoryless channel. Due to [44], [45], we know that $\lim_{n \to \infty} 1/n \log_2 M^*(n, p_e) = C$. Thus, the first order term in the finite-length expansion of $M^*(n, p_e)$ is nC. The higher order terms can be written as (see [5], [46])

$$\log_2 M^*(n, p_e) = nC - \sqrt{nV}Q^{-1}(p_e) + \mathcal{O}(\log n), \quad (3)$$

where V is the channel dispersion and $Q^{-1}(\cdot)$ is the inverse of the so called Q-function that is $Q(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-\frac{x^2}{2}} dx$. The third and higher order terms, however, depend on the particular channel under discussion (see [6, Theorem 41, 44, 73], [7]–[11]). More specifically, for BEC(ϵ), we have $C = 1 - \epsilon$, $V = \epsilon(1 - \epsilon)$ and the third order term is $\mathcal{O}(1)$. For BSC(δ), we have $C = 1 - h_2(\delta)$, $V = \delta(1 - \delta) \log_2^2((1 - \delta)/\delta)$, and the third order term is $1/2 \log_2 n + \mathcal{O}(1)$. For AWGN($\eta$), we have $C = 1/2 \log_2(1 + \eta)$, $V = \eta(\eta + 2)/(2(\eta + 1)^2 \ln^2 2)$, and the third order term is $1/2 \log_2 n + \mathcal{O}(1)$, as shown in [7].

The formula (3) is basically obtained by approximating the Random Coding Union (RCU) and converse bounds using Gaussian laws. This approach is best known for moderate-rate (or equivalently moderate-capacity) scenarios. There are,

---

[1] See [43] for an excellent review on this topic.
[2] In this paper, we only consider the average probability of error. Similar results can be obtained for the maximum probability of error.

however, other approaches in approximating the RCU and converse bounds which remain effective for a wider range of capacities such as the saddlepoint approximation [29]. Here, we briefly mention the formulation[3]. For a channel W with input distribution Q and a tuning factor $s > 0$, the information density is defined as

$$i_s(x, y) = \log \frac{W(y \mid x)^s}{\sum_{\bar{x}} Q(\bar{x}) W(y \mid \bar{x})^s} \ , \qquad (4)$$

where x and y represent the input and output of W. Then the saddlepoint approximation is

$$\widehat{\mathrm{rcu}}_s^*(n, M) = \beta_n(Q, R, s)\, e^{-n E_r(Q, R)}, \qquad (5)$$

where R is the rate, n is the blocklength. The error exponent $E_r(Q, R)$ is defined as

$$E_r(Q, R) = \sup_{s > 0, \rho \in [0,1]} -\log \mathbb{E}\left[e^{-\rho\, i_s(X, Y)}\right] - \rho R. \qquad (6)$$

Moreover, the coefficient $\beta_n(Q, R, s)$ is called the sub-exponential prefactor. The computation of $\beta_n(Q, R, s)$ is quite complicated and needs further steps. See [10] for the introduction and approximation of $\beta_n(Q, R, s)$ as well as a relevant analysis for $E_r(Q, R)$.

B. Polar Coding

In Section IV-B, we study state-of-the-art polar codes at low capacity. Polar codes were introduced by Arıkan in [41]. They are the first family of codes for the class of binary-input symmetric discrete memoryless channels that are provable to be capacity-achieving with low encoding and decoding complexity [41]. Polar codes and polarization phenomenon have been successfully applied to a wide range of problems including data compression [47], [48], broadcast channels [49], [50], multiple access channels [51], [52], physical layer security [53], [54], and coded modulations [55].

The basis of channel polarization consists of mapping two identical copies of the channel $W : \mathcal{X} \to \mathcal{Y}$ into the pair of channels $W^0 : \mathcal{X} \to \mathcal{Y}^2$ and $W^1 : \mathcal{X} \to \mathcal{X} \times \mathcal{Y}^2$, defined as

$$W^0(y_1, y_2 \mid x_1) = \sum_{x_2 \in \mathcal{X}} \frac{1}{2} W(y_1 \mid x_1 \oplus x_2) W(y_2 \mid x_2), \quad (7)$$

$$W^1(y_1, y_2, x_1 \mid x_2) = \frac{1}{2} W(y_1 \mid x_1 \oplus x_2) W(y_2 \mid x_2). \qquad (8)$$

Then, W^0 is a worse channel in the sense that it is degraded with respect to W, hence it is less reliable than W; and W^1 is a better channel in the sense that it is upgraded with respect to W, hence it is more reliable than W. The operation in (7) is also known as the *check* or *minus* operation and the operation in (8) is also known as the *variable* or *plus* operation.

By iterating this operation n times, we map $n = 2^m$ identical copies of the transmission channel W into the synthetic channels $\{W_m^{(i)}\}_{i \in \{0, \dots, n-1\}}$. More specifically, given $i \in \{0, \dots, n-1\}$, let $(b_1, b_2, \dots, b_m)$ be its binary expansion

<hr/>

[3]See [10], [30] for a thorough analysis of the saddlepoint approximation.

over m bits, where b_1 is the most significant bit and b_m is the least significant bit, i.e.,

$$i = \sum_{k=1}^m b_k 2^{m-k}. \qquad (9)$$

Then, we define the synthetic channels $\{W_m^{(i)}\}_{i \in \{0, \dots, n-1\}}$ as

$$W_m^{(i)} = (((W^{b_1})^{b_2})^{\cdots})^{b_m}. \qquad (10)$$

Example 1 (Synthetic Channel): Take $m = 4$ and $i = 10$. Then, the synthetic channel $W_4^{(10)} = (((W^1)^0)^1)^0$ is obtained by applying first (8), then (7), then (8), and finally (7).

The polar construction is polarizing in the sense that the synthetic channels tend to become either completely noiseless or completely noisy. Thus, in the encoding procedure, the k information bits are assigned to the positions (indices) corresponding to the best k synthetic channels. Here, the quality of a channel is measured by some reliability metric such as the Bhattacharyya parameter of the channel. The remaining positions are "frozen" to predefined values that are known at the decoder. As a result, the generator matrix of polar codes is based on choosing the k rows of the matrix $G_n = [1\ 0; 1\ 1]^{\otimes m}$ (with ";" separating the rows) which correspond to the best k synthetic channels.

III. Fundamental Limits

A. The Low-Capacity Regime

We first provide an informal description of the low-capacity regime and then proceed with a more formal specification. The low-capacity regime consists of two main components: (i) A channel W with capacity C. We think of C to be a very small number but *fixed*; (ii) The blocklength n which is defined as the number of times the channel W is used for transmission. Here, n should be thought of as a finite value, i.e., non-asymptotic.

We are interested in characterizing (optimal) ranges of n for which reliable transmission of a certain number of information bits is possible. Let k denote the number of information bits to be sent. To reliably communicate k bits, we clearly must have $n \geq k/C$ and thus n becomes fairly large when C is small. We are interested in finite-length values of n and their dependency on C and k. For example, assume that we aim to send a constant number of information bits k through the channel. We ask: What is the optimal (smallest) value of n to send k bits over the low-capacity channel with a given (fixed) error probability p_e? More precisely, we are searching over the set of all the possible values of n, such that $n \geq k/C$. In this search, we look for the smallest value of n for which reliable transmission of k bits with the desired error probability p_e is achievable.

One may ask why this question is practically relevant and/or worth a deeper theoretical investigation. We argue from two perspectives: (i) *Practical relevance:* There are many practical scenarios where the goal is to send a few bits over a low-capacity channel. For instance, in narrowband applications discussed in Section I, the number of information bits k is around a few tens, and the channel capacity C is typically

below 0.05. This makes n to vary in the range of a few thousand. For instance, if $k = 20$ and $C = 0.02$, then the blocklength n is at least 1000; (ii) *Fundamental limits:* As we will see, the low-capacity regime allows for simple and precise trade-offs between the length n and the number of information bits k, which depend on the channel and error probability. This stands in contrast to the case where C is not very small. For example, when $C = \frac{1}{2}$, sending 20 bits of information requires a fairly small blocklength n. In such a regime of n, it is intractable to provide precise closed-from estimates of the optimal blochlength and one typically resorts to (approximately) computing the well-known information-theoretic bounds, such as the random coding bound, among others. However, when C is small, the blocklength n becomes sufficiently large that allows us to provide simple, precise, and closed-form estimates of the optimal channel coding blocklength. Indeed, as we will see, such estimates require new techniques beyond the current methods used to analyze the finite-length limits of channel coding.

The low-capacity regime is not necessarily limited to transmitting a constant number of information bits, and one can consider other regimes of k. For example, assume that we would like to find the smallest value of n such that we can send $k = \alpha \log n$ bits of information reliably using n transmissions with an error probability not larger than p_e. How does the optimal (smallest) value of n scale with C? In the same manner, we may consider $k = \alpha\sqrt{n}$ and ask for the smallest value of n such that reliable transmission with k bits is possible. In this case, we clearly have $n \geq \alpha\sqrt{n}/C$, or equivalently $n \geq \alpha^2/C^2$. As a result, we are searching over the set of all the possible values of n, such that $n \geq \alpha^2/C^2$, and in this search, we look for the smallest value of n for which reliable transmission of $\alpha\sqrt{n}$ bits with a desired error probability p_e is achievable.

To proceed with a *formal* and *general* definition of the low-capacity regime, we need to provide a formal characterization of the term "low" in the finite-length regime where all the parameters such as C and n are assumed to be fixed and finite quantities. The low-capacity regime is formally defined using a channel W with capacity C and a function $f : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{n\to\infty} f(n)/n \to 0$. The main question is:

> **What is the smallest value of n such that $f(n)$ bits can be transmitted reliably in n transmissions over the channel W?**

A reliable communication necessitates $f(n)/n \leq C$ and $f(n) = nC$ indicates the maximum rate that is hypothetically achievable under any coding scheme. Define $\kappa = nC$. For the sake of analysis, one can treat n and C as variables and can hypothetically take them to the limits. As C gets smaller and smaller, we require n to get larger and larger for a reliable communication. In the limit, $C \to 0$ leads to $n \to \infty$ but the behavior of $\kappa = nC$ is a different story. In terms of n, κ can remain a constant or behave as a function, e.g., $\kappa = \log n$. This functionality is determined by the equality $f(n) = \kappa$. Hence, one can characterize the low capacity regime by determining the function f. As we will discuss later in this section, as n and C go to extremes, the expansion of $\log_2 M^*(n, p_e)$ in terms

of n (i.e., formula (3)) becomes less accurate. To tackle this, our approach is obtaining the expansion of $\log_2 M^*(n, p_e)$ in terms of κ rather than n. Suppose $\kappa = nC$ remains constant while $C \to 0$ and $n \to \infty$, then the expansion of $\log_2 M^*(n, p_e)$ in terms of κ remains stable despite n and C going to extremes. Every such expansion for $\log_2 M^*(n, p_e)$ in terms of κ remains valid when κ or equivalently $f(n)$ lies in a specific range of functions. As it can be seen in the following sections, the constraint over f only depends on its asymptotic behavior and thus the valid range for $f(n)$ (or equivalently κ) can be represented in terms of $\mathcal{O}(\cdot)$ notation.

Why the laws should be different in the low-capacity regime? In this paper, we investigate code design over channels with a very low capacity. Even though the formula (3) can still be used in the low-capacity regime, it provides a very loose approximation as (i) the channel variations in the low-capacity regime are governed by different probabilistic laws than the ones used to derive (3), and (ii) some of the terms hidden in $\mathcal{O}(\log n)$ will have significantly higher values and are comparable to the first and second terms. Similar arguments lead to the fact that the saddlepoint approximation (5) needs to be replaced with a more precise derivation in very low-capacity scenarios. Results provided in Section III will address these challenges.

Let us now explain why the current non-asymptotic laws of channel coding provided in (3) are not applicable in the low-capacity regime. Consider transmission over BEC(ϵ) with blocklength n. When the erasure probability ϵ is not too 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 (3) 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 [5], [6] and stated as [6, Corollary 42, Theorem 43] for BEC and [6, Corollary 39, Theorem 40] for BSC. These bounds are in a raw implicit form. The novel contribution of [5], [6] is in 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 [5], [6], i.e., we start with [6, Corollary 42, Theorem 43] for BEC and [6, 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 suitable for the low-capacity regime but not necessary for moderate values of C. These novel approaches in our analysis lead to the low-capacity coding bounds for BEC and BSC stated in the following subsections. For the case of an AWGN channel, it turns out that the bounds for the low-capacity regime are just a limiting case of the state-of-the-art bounds in the moderate-capacity regime. All Proofs are provided in the arXiv version of this paper [2].

B. The Binary Erasure Channel

As discussed earlier, the behavior of channel variations for the BEC in the low-capacity regime can be best approximated through the Poisson convergence theorem for rare events. This will lead to different (i.e., more accurate) non-asymptotic laws. Theorem 1 provides lower and upper bounds for the best achievable rate in terms of n, p_e, ϵ, and $\kappa := n(1-\epsilon)$. We use $\mathcal{P}_\lambda(x)$ to denote the Poisson cumulative distribution function with a slight modification that considers the probability of $X < x$ rather than $X \leq x$, i.e.,

$$\mathcal{P}_\lambda(x) = \Pr\{X < x\}, \qquad \text{where } X \sim \text{Poisson}(\lambda). \quad (11)$$

Theorem 1 (Finite-Length Coding Bounds for Low-Capacity BEC): Consider transmission over $\text{BEC}(\epsilon)$ and let $\kappa = n(1-\epsilon)$. Then, $M_1 \leq M^*(n, p_e) \leq M_2$, where M_1 is any (or the largest) value that satisfies

$$\mathfrak{P}_1(M_1) + 2\alpha_2\sqrt{\mathfrak{P}_1(M_1)} + \alpha_1\sqrt{\mathcal{P}_\kappa(\log_2 M_1)} - p_e \leq 0, \quad (12)$$

and M_2 is any (or the smallest) value that satisfies

$$\mathfrak{P}_2(M_2) - \alpha_2\sqrt{\mathfrak{P}_2(M_2)} - \alpha_1\sqrt{\mathcal{P}_\kappa(\log_2 M_2)} - p_e \geq 0, \quad (13)$$

and

$$\mathfrak{P}_1(M_1) = \mathcal{P}_\kappa(\log_2 M_1) + M_1 e^{-\kappa/2}\left(1 - \mathcal{P}_{\kappa/2}(\log_2 M_1)\right), \quad (14)$$

$$\mathfrak{P}_2(M_2) = \mathcal{P}_\kappa(\log_2 M_2) - \frac{e^\kappa}{M_2}\mathcal{P}_{2\kappa}\left(\log_2 M_2\right), \quad (15)$$

$$\alpha_1 = \frac{\sqrt{2}}{\epsilon^{3/2}}\left(\sqrt{e} - 1\right)(1 - \epsilon), \quad \alpha_2 = \frac{\sqrt{6}}{\epsilon^2\sqrt{\kappa}}\left(\sqrt{e} - 1\right)(1 - \epsilon). \quad (16)$$

Proof: See [2, Section A]. ∎

It is important to note that the analysis of Theorem 1 does not depend on the values of n and C and thus the results of Theorem 1 mathematically hold for all values of n, C, and p_e, i.e., for a moderate-capacity regime as well. However, they provide a sharp estimate when $\kappa = nC$ remains a moderate value despite a large n and a small C. Moreover, note that the bounds in Theorem 1 are expressed merely in terms of $\kappa := n(1 - \epsilon)$ rather than n. This agrees with

the intuition that the rate should depend on the amount of "information" passed through n usages of the channel rather than the number of channel uses n. Typically, the value of κ in low-capacity applications varies between a few tens to a few hundred. In such a range, no simple closed-form approximation of the Poisson distribution with mean κ exists. As a result, the lower and upper bounds in Theorem 1 cannot be simplified further. Furthermore, one can turn these bounds into bounds on the shortest (optimal) lengths n^* needed for transmitting k information bits with error probability p_e over a low-capacity BEC. In Section V, we numerically evaluate the lower and upper bounds predicted by Theorem 1 (see also [2, Section A.1]) and compare them with the prediction obtained from Formula (3) [5]. It is observed that our predictions are significantly more precise compared to the prediction obtained from Formula (3) and they become even more precise as the capacity approaches zero.

C. 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). To briefly explain the reason, consider transmission over $\text{BSC}(\delta)$ where the value of δ is close to 0.5. 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 $\delta \to 0.5$ one can write $\delta \approx 0.5 - \sqrt{\kappa/n}$ by using the Taylor expansion of the function $h_2(x)$ around $x = 0.5$. Transmission over $\text{BSC}(\delta)$ can be equivalently modeled as follows: (i) With probability 2δ, we let the output of the channel be chosen according to Bernoulli(0.5), 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). Given $\delta \to 0.5$, 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 (3) 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 (Finite-Length Coding Bounds for Low-Capacity BSC): Consider transmission over $\text{BSC}(\delta)$ in low-capacity regime in the sense that the function $f(n)$ mentioned in

Section III-A belongs to the class $(o(n))^{2/3}$ and let $\kappa = n(1 - h_2(\delta))$. Then,

$$\log_2 M^*(n, p_e) = \kappa - 2\sqrt{\frac{2\kappa\delta(1-\delta)}{\ln 2}} Q^{-1}(p_e) + \frac{1}{2}\log_2 \kappa \\ - \log_2 p_e + \mathcal{O}(\log\log\kappa). \quad (17)$$

Proof: See [2, Section B.1]. ∎

Following the discussion of Section III-A, note that the low-capacity regime considered in Theorem 2 is specified by the function f belonging to the class $(o(n))^{2/3}$. This means for the estimate (17) to hold, we must have $\kappa = f(n) = (o(n))^{2/3}$. This constraint comes from using the condition $\kappa\sqrt{\kappa} = o(n)$ in the proof of Theorem 2. Moreover, we remark that the $\mathcal{O}(\log\log\kappa)$ term contains some other terms such as $\mathcal{O}(\sqrt{-\log p_e}/\log\kappa)$. For practical scenarios, the term $\mathcal{O}(\log\log\kappa)$ will be dominant.[4] We also note that similar to the BEC case, all terms in (17) are expressed in terms of κ rather than n. This agrees with the intuition that the rate should depend on the amount of "information" passed through n usages of the channel rather than n itself.

Corollary 1: Consider transmission of k information bits over a low-capacity BSC(δ) that is specified in Theorem 2. Then, the optimal blocklength n^* for such transmission is

$$n^* = \frac{1}{1 - h_2(\delta)}\left(k + 2\sqrt{\frac{2\kappa\delta(1-\delta)}{\ln 2}}Q^{-1}(p_e) \right. \\ \left. + \frac{4\delta(1-\delta)}{\ln 2}Q^{-1}(p_e)^2 + \log_2 p_e + \mathcal{O}(\log k)\right). \quad (18)$$

Proof: See [2, Section B.2]. ∎

D. The Additive White Gaussian Noise Channel

First, let us further clarify our description of coding over the AWGN channel. We consider n uses of a real AWGN channel in which the input X_i and the output Y_i at each $i = 1, \ldots, n$ are related as $Y_i = X_i + Z_i$. Here, the noise term $\{Z_i\}_{i=1}^n$ is a memoryless, stationary Gaussian process with zero mean and unit variance. Given an (M, p_e)-code for W^n, where W is the AWGN channel, a cost constraint on the codewords must be applied. The most commonly used cost is

$$\forall m \in \mathcal{M}: \quad \|f_{enc}(m)\|_2^2 = \sum_{i=1}^n (f_{enc}(m))_i^2 \le \eta\, n, \quad (19)$$

where η, with a slight abuse of notation, refers to SNR. Since characterization of the code depends on the SNR η, we denote an (M, p_e)-code and $M^*(n, p_e)$ by (M, p_e, η)-code and $M^*(n, p_e, \eta)$, respectively.

Similar to BSC, the channel variations in low-capacity AWGN channels are best approximated by the central limit theorem. The following theorem is obtained by using the ideas in [6, Theorem 73] with slight modifications. It turns out that coding bounds for AWGN in the low-capacity regime can be obtained as a limiting case of the state-of-the-art bounds in the moderate-capacity regime. The following theorem and corollary are resulted simply by repeating the same argument

[4]We include only the dominant term inside $\mathcal{O}(\cdot)$.

in a manner that remains valid for these limiting cases. This needs a tiny refinement of the analysis which is done in the corresponding proofs, provided in the arXiv version of this paper [2].

Theorem 3 (Finite-Length Coding Bounds for Low-Capacity AWGN): Consider transmission over AWGN(η) in low-capacity regime and let $\kappa = \frac{n}{2}\log_2(1+\eta)$. Then,

$$\log_2 M^*(n, p_e, \eta) = \kappa - \frac{\sqrt{\eta+2}}{(\eta+1)\sqrt{\ln 2}} \cdot \sqrt{\kappa}\, Q^{-1}(p_e) + \mathcal{E}, \quad (20)$$

where

$$\mathcal{O}(1) \le \mathcal{E} \le \frac{1}{2}\log_2\kappa - \log_2 p_e + \mathcal{O}\left(\frac{1}{\sqrt{-\log p_e}}\right). \quad (21)$$

Proof: See [2, Section C.1]. ∎

The same considerations about $\mathcal{O}(\cdot)$ notation, as discussed earlier, should be taken into account here. Also note that as for BEC and BSC, the optimal blocklength for the AWGN channel can be expressed in terms of other parameters in the low-capacity regime which is stated in the following corollary.

Corollary 2: Consider transmission of k information bits over a low-capacity AWGN(η). Then, the optimal blocklength n^* for such transmission is

$$n^* = \frac{2}{\log_2(1+\eta)}\left(k + \frac{\sqrt{\eta+2}}{(\eta+1)\sqrt{\ln 2}}Q^{-1}(p_e)\cdot\sqrt{k} \right. \\ \left. + \mathcal{O}\left(\log\frac{1}{p_e}\right)\right). \quad (22)$$

Proof: See [2, Section C.2]. ∎

As we mentioned earlier, Theorem 3 is obtained directly by taking the limit of the AWGN results in [5]. The Corollary 2 is consequently a limiting case of the moderate-regime analysis in [5]. Therefore, the contribution of these results is merely the emphasis on the fact that when the underlying channel is AWGN, the moderate-capacity regime analysis in [5] may extend well to the low-capacity regime only by taking the limit. Having this, we do not numerically evaluate Corollary 2 on the AWGN channel.

IV. THE ANALYSIS OF PRACTICAL CODE DESIGNS

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 low-density parity-check (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 a poor performance. E.g., an (l, r)-regular LDPC code with design rate $R = 0.01$ requires $r, l \ge 99$. Hence, the Tanner graph will have a minimum degree of at least 99 and even for code block lengths on the order of tens of thousands, the Tanner graph will have many short cycles. To circumvent this issue, the current practical designs, e.g., the NB-IoT standard, use repetition coding. More specifically, a low-rate repetition code is concatenated with a powerful moderate-rate code. For example, an iterative code of rate R and length n/r can be repeated r times to construct a code of length n with rate R/r. In Section IV-A, we will discuss

the advantages and drawbacks of using repetition schemes along with trade-offs between the number of repetitions and the performance of the code.

Unlike iterative codes, polar codes and most algebraic codes (e.g., BCH or Reed-Muller codes) can be used without any modification for low-rate applications. In Section IV-B, we look into polar coding for low-capacity channels. In particular, we show that polar coding is advantageous in terms of inherently adopting an optimal number of repetitions. Theorems 4 and 5 provide tight bounds on the optimal number of repetitions in terms of the capacity, that are not specific to polar codes, and Theorem 6 shows that the construction of polar codes naturally adopts a certain number of repetition blocks in the low-capacity regime that match the optimal number of repetitions (up to a constant multiplicative factor).

Throughout this section, we consider code design for the class of binary memoryless symmetric (BMS) channels. A BMS channel W has binary input and, letting $W(y \mid x)$ denote the transition matrix, there exists a permutation π on the output alphabet such that $W(y \mid 0) = W(\pi(y) \mid 1)$. Notable exemplars of this class are BEC, BSC, and BAWGN channels.[5]

A. How Much Repetition is Needed?

As mentioned earlier, repetition is a straightforward way to design practical low-rate codes while utilizing the power of state-of-the-art code designs. Let r be a divisor of n, where n denotes the length of the code. Repetition coding consists of designing first a smaller outer code of length n/r and then repeating each of the coded bits r times (i.e., the inner code is a repetition code of rate $1/r$). The length of the overall code is then $n/r \cdot r = n$. This is equivalent to transmitting the outer code over the *r-repetition channel* W^r which takes a bit as the input and outputs an r-tuple which 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 = $ BEC(ϵ^r).

The main advantage of repetition coding is the reduction in computational complexity, especially when r is large. This is because the encoding/decoding complexity is effectively reduced to that of the outer code, i.e., once the outer code is constructed, at the encoding side, we just need to repeat each of its coded bits r times, and at the decoding side the log-likelihood of an r-tuple consisting of r independent transmissions of a bit is equal to the sum of the log-likelihoods of the individual channel outcomes. The computational latency of the encoding and decoding algorithms is reduced to that of the outer code in a similar fashion.

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 $C(W^r) \leq rC(W)$ 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 the largest repetition size r_β is such that

$$\frac{n}{r_\beta} C(W^{r_\beta}) \geq \beta n C(W). \tag{23}$$

Let us first assume that transmission takes place over BEC(ϵ). We thus have $W^r = $ BEC(ϵ^r). If ϵ is not close to 1, then even $r = 2$ would result in a considerable rate loss, e.g., if $\epsilon = 0.5$, then $C(W^2) = 0.75$ whereas $2C(W) = 1$. However, when ϵ is close to 1, then at least for small values of r the rate loss can be negligible, e.g., for $r = 2$, we have $C(W^2) = 1 - \epsilon^2 \approx 2(1 - \epsilon) = 2C(W)$. The following theorem provides lower and upper bounds for the largest repetition size r_β that satisfies (23).

Theorem 4 (Maximum Repetition Length for BEC): If $W = $ BEC(ϵ), then for the largest repetition size r_β that satisfies (23), we have

$$\frac{n(1-\epsilon)\ell}{2\left(1 - \frac{\beta}{\ell}\right)} \cdot \left(\frac{\beta}{\ell}\right)^2 \leq \frac{n}{r_\beta} \leq \frac{n(1-\epsilon)\ell}{2\left(1 - \frac{\beta}{\ell}\right)}, \tag{24}$$

where $\ell = -(\ln \epsilon)/(1 - \epsilon)$. Equivalently, assuming $\kappa = n(1 - \epsilon)$, (24) becomes

$$\frac{\kappa}{2(1-\beta)} \cdot \beta^2(1 + \mathcal{O}(1 - \epsilon)) \leq \frac{n}{r_\beta} \leq \frac{\kappa}{2(1-\beta)}(1 + \mathcal{O}(1 - \epsilon)). \tag{25}$$

Proof: See [2, Section D.1]. ■

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

$$\frac{n}{r_\beta} C(W^{r_\beta}) = nC(W) + \mathcal{O}(1). \tag{26}$$

If $W = $ BEC(ϵ) and $\kappa = n(1-\epsilon)$, then the necessary condition implies plugging $\beta = 1 - \mathcal{O}(1/\kappa)$ into (23). Moreover, from Theorem 4, 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 $\mathcal{O}(\kappa^2)$.

A noteworthy conclusion from the above remark is that as having negligible rate loss implies the repetition size to be at most $\mathcal{O}(n/\kappa^2)$, then 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 a few tens. Thus, the idea of using codes such as 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 a non-negligible rate loss. In contrast, the polar coding construction has implicitly a repetition block of optimal size $\mathcal{O}(n/\kappa^2)$ as we will see in the next section.

In [2, Section D.2], we prove the following theorem stating that the binary erasure channel has the smallest rate loss due

---

[5]For a relevant study of code design at a low-SNR scenario, see [56], [57].

to repetition among all the BMS channels. This provides an upper bound on r_β for any BMS channel.

Theorem 5 (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 (23). Hence, for any BMS channel with capacity C and $\kappa = nC$, we have

$$\frac{n}{r_\beta} \geq \frac{\kappa}{2(1-\beta)}\beta^2(1 + \mathcal{O}(C)). \tag{27}$$

Proof: See [2, Section D.2]. ∎

Remark 2: Similar to Remark 1, we can conclude that for any BMS channel with low capacity, 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 and Repetition at Low Capacity

We have shown in Section IV-A that the maximum allowed repetition size to have negligible capacity loss is $\mathcal{O}(n/\kappa^2)$. We show in this section that at low-capacity regime, 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 naturally adopted repetitions. That is, the encoding complexity can be reduced to $n + \mathcal{O}(\kappa^2 \log \kappa)$ and the decoding complexity using the successive cancellation list (SCL) decoder with list size L, proposed by Tal and Vardy [58], is reduced to $n + \mathcal{O}(L\kappa^2 \log \kappa)$. Recall that the original implementation of polar codes requires $\mathcal{O}(n \log n)$ encoding complexity and $\mathcal{O}(Ln \log n)$ decoding complexity. Moreover, as the operations involving repeated blocks 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, the simplified SC decoder [59] or relaxed polar codes [60] can be invoked. Such complexity reductions are important for making polar codes a suitable candidate in practice. In a related work, designing low-rate codes for BSCs by concatenating high rate polar codes together with repetitions is considered [61]. Furthermore, following the publication of the initial version of this paper, several works have looked into improving low-rate polar codes by either tweaking the repetition or concatenating them with other codes [62]–[64].

Theorem 6: 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 $\frac{1}{2}$ has at least $m - m_0$ 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.

Proof: See [2, Section D.3]. ∎

Remark 3: Note that from Theorem 6, 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 1 and 2.

V. NUMERICAL ANALYSIS OF FUNDAMENTAL LIMITS

In this section, we numerically evaluate our channel coding bounds from Section III. We report the numerical results on the BEC and the BSC cases. As mentioned earlier, we do not numerically implement our results on the AWGN channel (Theorem 3 and Corollary 2) since their contribution is merely showing that the limit of the existing moderate-capacity regime analysis will lead to a low-capacity regime bound.

For the BEC, we have compared in Figure **1**, the lower and upper bounds obtained from Theorem 1 with the predictions of Formula (3). We have also plotted the performance of polar codes. 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 $p_e = 0.01$. 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 that satisfies equation (13) with $\kappa = n(1-\epsilon)$ and $p_e = 0.01$.
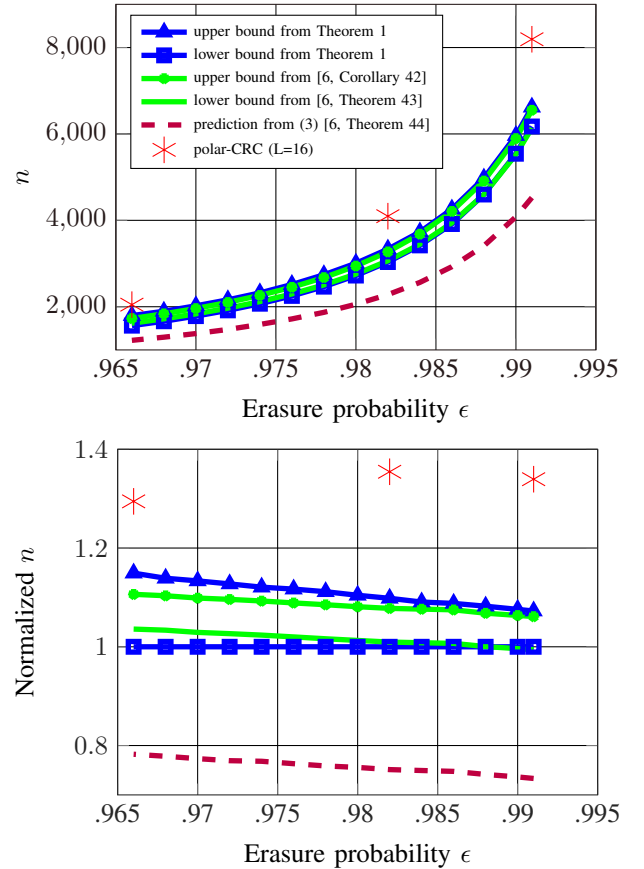

Figure 1. Comparison for low-capacity BECs. The number of information bits is $k = 40$ and the target error probability is $p_e = 0.01$. For the upper plot, with the same legend entries as the lower plot, all the blocklengths n in the lower plot are normalized by the value of the lower bound, obtained from Theorem 1.

Note that [6, Corollary 42] and [6, Theorem 43], also presented in [2, Section A], provide the raw upper and lower bounds for the optimal blocklength in BEC. Both the classical

estimation (3) (whose precise version for BEC can be found in [6, Theorem 44]) and Theorem 1 are estimating these raw bounds. As expected, the prediction obtained from [6, Theorem 44] is not precise in the low-capacity regime and it becomes worse as the capacity approaches zero. On the other hand, the estimated bounds from Theorem 1 converge to the original raw bounds of [6, Corollary 42] and [6, Theorem 43] as the capacity approaches zero. Also, the performance of the polar code is shown in Figure **1**. The polar code is concatenated with cyclic redundancy check (CRC) code of length 6 and is decoded with the list-SC algorithm [58] with list size $L = 16$.

Figure **2** considers the scenario of sending $k = 40$ bits of information over a low-capacity BSC with target error probability $p_e = 0.01$. We have compared in Figure **2**, the predictions from Theorem 2 and Formula (3) (we used a precise version of Formula (3) for BSC given in [6, Theorem 41]) together with the raw upper and lower bounds from [6, Corollary 39] and [6, Theorem 40] that are also presented in [2, Section B]. Note that both Theorem 2 and [6, Theorem 41] provide single predictions to estimate the true value of the optimal blocklength that lies between the aforementioned raw bounds. Therefore, it is not abnormal if neither of the predictions from [6, Theorem 41] or Theorem 2 lie between the raw bounds. In this way, Figure **2** shows that, as we expected, the prediction from [6, Theorem 41] is quite imprecise in the low-capacity regime, particularly in comparison to the prediction from Theorem 2 which is exact up to $\mathcal{O}(\log\log\kappa)$ terms. The performance of polar codes is also plotted in Figure **2**. An interesting problem is to analyze the finite-length scaling of polar codes in the low-capacity regime [65]–[69].

Figure **3** represents the same setting as above, i.e., sending $k = 40$ information bits over the BEC(ϵ) with the desired probability of error $p_e = 0.01$ for BEC (up) and BSC (down). The vertical axis in both cases is the normalized blocklength and the horizontal axis is the erasure probability ϵ and the crossover probability δ for the corresponding plots. In both plots, the corresponding channel parameter varies in an extremely low-capacity region. In such a region, we plotted the RCU achievability (upper) and converse (lower) bounds, our approximation from Theorem 1 (for BEC) and Theorem 2 (for BSC), the saddlepoint approximation [10], [30], and a simplified version of it known as the exact asymptotics which we simply refer to as the "simplified saddlepoint approximation" to avoid confusion. Since the simplified version of the saddlepoint approximation is by construction less precise than the original version, we focus only on the comparison of our results to the original saddlepoint approximation.

In the BEC case, i.e., Figure **3** (up), it can be seen that our approximations are dramatically tight and converge to the true bounds as the capacity goes to zero. However, the saddlepoint approximation does not perform well enough in this regime. This behavior can be described by noticing that despite the fact that the saddlepoint approximation uses an innovative approach to estimate the RCU bounds in all rates, it is still based on Gaussian laws; however, the channel coding bounds for BEC, as this paper shows, are best described by
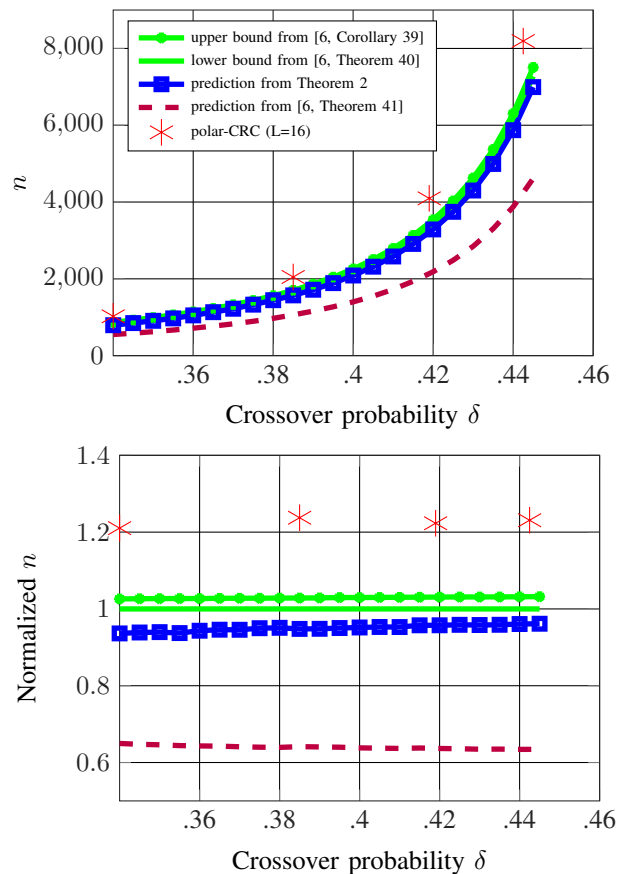

Figure 2. Comparison for low-capacity BSC. The number of information bits is $k = 40$ and the target error probability is $p_e = 0.01$. For the upper plot, with the same legend entries as the lower plot, all the blocklengths n in the lower plot are normalized by lower bound given by [6, Theorem 40].

the Poisson laws rather than Gaussian laws in this regime.

In the BSC case where we provide a single prediction of the optimal blocklength that satisfies both RCU achievability and converse bounds, Figures **3** shows that this prediction remains close and slowly converges to the true bounds as the capacity decreases. On the other hand, although the saddlepoint approximation works well when the capacity is not small enough, it starts to break down in the region where the capacity becomes extremely small. Note that here both saddlepoint and our method enjoy the Gaussian laws as the baseline of the analysis, however, the weight pattern of the terms in the true bounds and the saddlepoint approximation starts to deviate as the capacity gets closer to zero and thus the approximation starts to lose precision.

VI. Conclusion and Future Work

In this paper, we specified a notion of the low-capacity regime for channel coding and studied channel coding at such regimes from two major perspectives, namely, finite-length fundamental limits and code constructions. More specifically, finite-length analysis specific to the low-capacity regime was carried out for several types of channels including binary erasure channels (BECs), binary symmetric channels (BSCs), and additive white Gaussian noise (AWGN) channels. Furthermore, in the context of code construction, the optimal
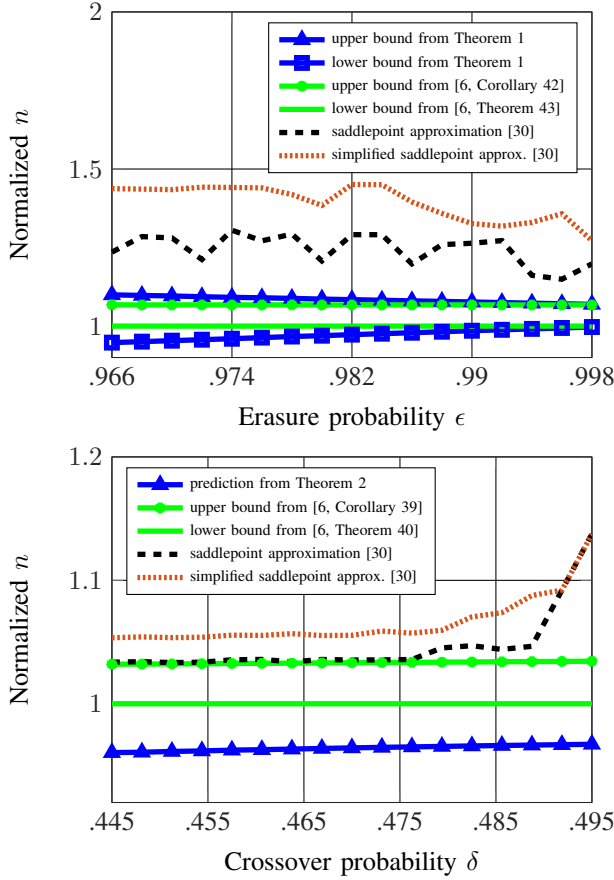
Figure 3. Sending $k = 40$ information bits over BEC (up) and BSC (down) in the low-capacity regime for the target error probability $p_e = 0.01$ to compare our approximations from Theorem 1 and Theorem 2 or equivalently Corollary 1 with saddlepoint approximations [30].

number of repetitions was characterized for transmission over binary memoryless symmetric (BMS) channels, in terms of the code block length and the underlying channel capacity. It was further shown that capacity-achieving polar codes naturally adopt the aforementioned optimal number of repetitions.

There are several directions for future work. In terms of fundamental limits, it is interesting to study different classes of discrete memoryless channels beyond the ones considered in this paper to characterize their fundamental non-asymptotic laws of channel coding in the low-capacity regime. In terms of code constructions, it is important to study concatenation schemes with low-complexity decoding algorithms comparable to those of straightforward repetition schemes, which can potentially lead to higher rates. From a practical implementation perspective, there are various other challenges besides channel coding in order to enable reliable communications at very low-rate regimes. This includes detection, synchronization, and multi-user communication. For instance, synchronization requires transmission of certain pilot signals which would be difficult to detect due to having a very low power. Alternatively, one can explore non-coherent communications at very low rates which is another interesting research direction. Studying low-rate communications together with multi-user schemes is another important problem. This

becomes relevant especially for IoT applications where a massive number of low-power users are present in the field. To this end, various approaches including grant-free and uncoordinated multiple access [70]–[72] as well as scalable coded non-orthogonal techniques [73] can be explored.

REFERENCES

[1] M. Fereydounian, M. V. Jamali, H. Hassani, and H. Mahdavifar, "Channel coding at low capacity," in *Proc. 2019 IEEE Inf. Theory Workshop (ITW)*, Aug 2019, pp. 1–5.

[2] M. Fereydounian, H. Hassani, M. V. Jamali, and H. Mahdavifar, "Channel coding at low capacity," *arXiv preprint arXiv:1811.04322*, 2023.

[3] R. Dobrushin, "Mathematical problems in the Shannon theory of optimal coding of information," in *Proc. 4th Berkeley Symp. Math., Statist., and Probability*, vol. 1, 1961, pp. 211–252.

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

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

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

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

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

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

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

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

[12] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Transactions on communications*, vol. 44, no. 10, pp. 1261–1271, 1996.

[13] R. Gallager, "Low-density parity-check codes," *IRE Trans. on Info. Theory*, vol. 8, no. 1, pp. 21–28, January 1962.

[14] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. on Info. Theory*, vol. 45, no. 2, pp. 399–431, March 1999.

[15] 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. (CSCN)*, Berlin, Germany, Oct./Nov. 2016, pp. 1–7.

[16] R. Ratasuk, B. Vejlgaard, N. Mangalvedhe, and A. Ghosh, "NB-IoT system for M2M communication," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2016, pp. 1–5.

[17] K. S. Andrews, D. Divsalar, S. Dolinar, J. Hamkins, C. R. Jones, and F. Pollara, "The development of turbo and LDPC codes for deep-space applications," *Proceedings of the IEEE*, vol. 95, no. 11, pp. 2142–2156, 2007.

[18] H. Liang, A. Liu, X. Tong, and C. Gong, "Raptor-like rateless spinal codes using outer systematic polar codes for reliable deep space communications," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2020, pp. 1045–1050.

[19] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Block-fading channels at finite blocklength," in *ISWCS 2013; The Tenth International Symposium on Wireless Communication Systems*. VDE, 2013, pp. 1–4.

[20] W. Yang, *Fading channels: Capacity and channel coding rate in the finite-blocklength regime*. Chalmers University of Technology, 2015.

[21] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multipleantenna fading channels at finite blocklength," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4232–4265, 2014.

[22] Y.-W. Huang and P. Moulin, "Finite blocklength coding for multiple access channels," in *2012 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2012, pp. 831–835.

[23] V. Y. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 881–903, 2013.

[24] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, 2012.

[25] R. Nomura and T. S. Han, "Second-order slepian-wolf coding theorems for non-mixed and mixed sources," *IEEE transactions on information theory*, vol. 60, no. 9, pp. 5553–5572, 2014.

[26] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3493–3503, 2016.

[27] M. Tahmasbi and M. R. Bloch, "First-and second-order asymptotics in covert communication," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2190–2212, 2018.

[28] S. Watanabe, S. Kuzuoka, and V. Y. Tan, "Nonasymptotic and second-order achievability bounds for coding with side-information," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1574–1605, 2015.

[29] A. Martinez and A. G. i Fabregas, "Saddlepoint approximation of random-coding bounds," in *2011 Information Theory and Applications Workshop*. IEEE, 2011, pp. 1–6.

[30] J. Scarlett, A. Martinez, and A. G. i Fàbregas, "The saddlepoint approximation: Unified random coding asymptotics for fixed and varying rates," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 1892–1896.

[31] J. Font-Segura, G. Vazquez-Vilar, A. Martinez, A. Guillén i Fàbregas, and A. Lancho, "Saddlepoint approximations of lower and upper bounds to the error probability in channel coding," in *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, 2018, pp. 1–6.

[32] Y. Altuğ and A. B. Wagner, "On exact asymptotics of the error probability in channel coding: symmetric channels," *IEEE Transactions on Information Theory*, 2020.

[33] J. Honda, "Exact asymptotics of random coding error probability for general memoryless channels," in *2018 IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 1844–1848.

[34] B. Reiffen, "A note on "very noisy" channels," *Information and Control*, vol. 6, no. 2, pp. 126–130, 1963.

[35] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 1, pp. 3–18, 1965.

[36] E. E. Majani, "A model for the study of very noisy channels, and applications," Ph.D. dissertation, California Institute of Technology, 1988.

[37] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel—Parts I and II," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1449–1471, 1988.

[38] Y. Sakai, V. Y. Tan, and M. Kovačević, "Second-and third-order asymptotics of the continuous-time Poisson channel," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 4742–4760, 2020.

[39] A. B. Wagner, N. V. Shende, and Y. Altuğ, "A new method for employing feedback to improve coding performance," *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 6660–6681, 2020.

[40] N. V. Shende and A. B. Wagner, "On very noisy channels with feedback," in *Proc. 55th Annu. Allerton Conf. Commun., Control, and Comput.* IEEE, 2017, pp. 852–859.

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

[42] M. Talagrand, "The missing factor in Hoeffding's inequalities," *Annales de l'I.H.P. Probabilités et statistiques*, vol. 31, no. 4, pp. 689–702, 1995.

[43] V. Y. Tan *et al.*, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations and Trends® in Communications and Information Theory*, vol. 11, no. 1-2, pp. 1–184, 2014.

[44] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech. Journal*, no. 27, pp. 379–423, 1948.

[45] J. Wolfowitz, "Notes on a general strong converse," *Inf. Contr.*, vol. 12, pp. 1–4, 1968.

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

[47] E. Arıkan, "Source polarization," in *Proc. 2010 IEEE Int. Symp. Inf. Theory (ISIT)*, 2010, pp. 899–903.

[48] E. Abbe, "Polarization and randomness extraction," in *Proc. 2011 IEEE Int. Symp. Inf. Theory (ISIT)*, 2011, pp. 184–188.

[49] M. Mondelli, S. H. Hassani, I. Sason, and R. L. Urbanke, "Achieving Marton's region for broadcast channels using polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 783–800, 2015.

[50] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for broadcast channels," *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 758–782, 2015.

[51] E. Şaşoğlu, E. Telatar, and E. Yeh, "Polar codes for the two-user binary-input multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6583–6592, 2013.

[52] H. Mahdavifar, M. El-Khamy, J. Lee, and I. Kang, "Achieving the uniform rate region of general multiple access channels by polar coding," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 467–478, 2016.

[53] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6428–6443, 2011.

[54] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, "Nested polar codes for wiretap and relay channels," *IEEE Commun. Lett.*, vol. 14, no. 8, pp. 752–754, 2010.

[55] H. Mahdavifar, M. El-Khamy, J. Lee, and I. Kang, "Polar coding for bit-interleaved coded modulation," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3115–3127, 2015.

[56] E. C. Posner, "Properties of error-correcting codes at low signal-to-noise ratios," *SIAM J. Appl. Math.*, vol. 15, no. 4, pp. 775–798, 1967.

[57] Chi-Chao Chao, R. J. McEliece, L. Swanson, and E. R. Rodemich, "Performance of binary block codes at low signal-to-noise ratios," *IEEE Trans. Inf. Theory*, vol. 38, no. 6, pp. 1677–1687, Nov 1992.

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

[59] A. Alamdar-Yazdi and F. R. Kschischang, "A simplified successive-cancellation decoder for polar codes," *IEEE Commun. Lett.*, vol. 15, no. 12, pp. 1378–1380, 2011.

[60] M. El-Khamy, H. Mahdavifar, G. Feygin, J. Lee, and I. Kang, "Relaxed polar codes," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 1986–2000, 2017.

[61] I. Dumer, "Polar codes with a stepped boundary," in *Proc. 2017 IEEE Int. Symp. Inf. Theory (ISIT)*. IEEE, 2017, pp. 2613–2617.

[62] F. Abbasi, H. Mahdavifar, and E. Viterbo, "Polar coded repetition," *IEEE Transactions on Communications*, vol. 70, no. 10, pp. 6399–6409, 2022.

[63] I. Dumer and N. Gharavi, "Codes approaching the shannon limit with polynomial complexity per information bit," in *2021 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2021, pp. 238–243.

[64] F. Abbasi, H. Mahdavifar, and E. Viterbo, "Hybrid non-binary repeated polar codes," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7582–7594, 2022.

[65] S. H. Hassani, K. Alishahi, and R. L. Urbanke, "Finite-length scaling for polar codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5875–5898, 2014.

[66] M. Mondelli, S. H. Hassani, and R. L. Urbanke, "Unified scaling of polar codes: Error exponent, scaling exponent, moderate deviations, and error floors," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6698–6712, 2016.

[67] D. Goldin and D. Burshtein, "Improved bounds on the finite length scaling of polar codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6966–6978, 2014.

[68] V. Guruswami and P. Xia, "Polar codes: Speed of polarization and polynomial gap to capacity," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 3–16, 2015.

[69] H. Mahdavifar, "Polar coding for non-stationary channels," *IEEE Transactions on Information Theory*, vol. 66, no. 11, pp. 6920–6938, 2020.

[70] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. De Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the internet of things," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 88–99, 2018.

[71] Y. Polyanskiy, "A perspective on massive random-access," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 2523–2527.

[72] V. K. Amalladinne, J.-F. Chamberland, and K. R. Narayanan, "A coded compressed sensing scheme for unsourced multiple access," *IEEE Trans. Inf. Theory*, vol. 66, no. 10, pp. 6509–6533, 2020.

[73] M. V. Jamali and H. Mahdavifar, "Massive coded-NOMA for low-capacity channels: A low-complexity recursive approach," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3664–3681, 2021.