# Channel Coding at Low Capacity

Mohammad Fereydounian
University of Pennsylvania
3401 Walnut St, Philadelphia, PA 19104
mferey@seas.upenn.edu

Mohammad Vahid Jamali
University of Michigan
1301 Beal Avenue, Ann Arbor, MI 48109
mvjamali@umich.edu

Hamed Hassani
University of Pennsylvania
3401 Walnut St, Philadelphia, PA 19104
hassani@seas.upenn.edu

Hessam Mahdavifar
University of Michigan
1301 Beal Avenue, Ann Arbor, MI 48109
hessam@umich.edu

**Abstract**

Low-capacity scenarios have become increasingly important in the technology of Internet of Things (IoT) and next generation of mobile networks. Such scenarios require efficient, reliable transmission of information over channels with 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.

## 1 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. Consequently, it is expected that the total number of IoT devices supported through cellular networks will reach 1.5 billion by 2021 [1]. To ensure high coverage, the standard has to support coupling loss as large as 170 dB for these applications, which is approximately 20 dB higher than that of the legacy LTE. As stated in [1,3], tolerating such coupling loss requires reliable detection for a typical $-13$ dB of effective SNR, translated to capacity $\approx 0.03$. 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 $1.6 \times 10^{-4}$ [1]. However, from a channel coding perspective, repeating a high-rate code to enable low-rate communication can be very sub-optimal.

Surprisingly, a similar situation arises in wideband scenarios, and in particular in the mmWave band. In the simplest model for a wideband channel, transmission takes place over an additive white Gaussian noise (AWGN) channel with the capacity $C = B \log(1 + \frac{P}{N_0 B})$, where $P$ is the signal power, $N_0$ is the noise variance, and $B$ is the allocated bandwidth. Assuming a limited transmission power $P$ and high bandwidth $B \to \infty$, we operate in the low-capacity regime–in terms of the underlying channel code rate per symbol. Note that this is not in contrast with the high data rates, in terms of bits per second, of wideband applications. In other words, as $B$ grows large, we are able to transmit a growing number of symbols in a fixed time interval. However, these symbols experience a vanishing SNR when the total power is fixed. More specifically, we have $SNR \to 0$ as $B \to \infty$.

Most of classical channel coding theory is centered on the designs 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 [4,5], and a precise characterization is provided in [6] 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 a characteristic of the channel referred to as channel dispersion. Such non-asymptotic laws have steered optimal code design for typical channels,. However, very little is known about optimal code design in the low-capacity regime where the capacity of the channel $C$ could be as small as $\mathcal{O}(1/n)$ and hence the first and second term of the law could be as small as the third terms (i.e., the $o(1)$ term). The low-capacity regime consists of sending $k$ bits of information, where $k$ could be as small as few tens, over a channel with very low capacity, e.g., $C \leq 0.01$. To communicate reliably in this regime, we require codes with very large length $n$ albeit the fact that the overall capacity of $n$ channel usages, $nC$, could be small. Indeed, optimal code design in the low-capacity regime 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($\epsilon$) with blocklength $n$. When the erasure probability $\epsilon$ 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 rate. 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 [7] can naturally be adapted to the low-rate regime, however, the current implementation of these codes suffers from relatively high computational complexity and latency. Note that there is a subtle difference between the low-rate regime and the moderate-rate regime when characterizing the behavior of complexity and latency of decoders. These parameters are often described as functions of code blocklength $n$ and, in the moderate-rate regime, result in the same expression if we replace $n$ by the number of information bits $k$ which scales linearly with $n$. However, this does not necessarily hold for the low-rate regime as $k$ is significantly smaller than $n$. For instance, the decoding complexity and latency of polar codes are known to be $\mathcal{O}(n \log n)$ and $\mathcal{O}(n)$, respectively [7]. While this is reasonable when $k$ scales linearly with $n$, it becomes inefficient when $k$ is a sub-linear function of $n$. We essentially need low-latency decoders, in terms of $k$, in order to provide high data rates, in terms of bits per second (b/s), in wideband applications. We also need low-complexity decoders to provide low device unit cost and low power consumption for narrowband applications. In practice, the proposed solution for NB-IoT code design is simply to apply many repetitions on an underlying code such as a Turbo code or a polar code. Even though this approach leads to efficient implementations, the rate loss through many repetitions will result in codes with mediocre performance.

This paper provides the first comprehensive study of channel coding in the low-capacity regime. In Section 2, we will provide the necessary background. In Section 3, we will formally define the low-capacity regime and provide fundamental non-asymptotic laws of channel coding for a diverse set of channels with practical significance: the binary erasure channel, the binary symmetric channel, and the additive white Gaussian channel. Section 4 considers various approaches to practical code design in the low-capacity regime with numerical comparisons with the non-asymptotic bounds derived in Section 3 as well as the codes used in the NB-IoT standard.

## 2 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 $p_e$, that is[2]

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

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 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

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

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

- BEC($\epsilon$): binary erasure channel with erasure probability $\epsilon$.
- BSC($\delta$): binary symmetric channel with crossover probability $\delta$.
- AWGN($\eta$): additive white Gaussian noise channel with signal-to-noise ratio (SNR) $\eta$.

[1]For more details, we refer the reader to [8] 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 maximum probability of error.

Let us further clarify our description of coding over the AWGN channel. We consider $n$ uses of the 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

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

where $\eta$ is the SNR. Since, characterization of the code depends on the SNR $\eta$, we denote an $(M, p_e)$-code and $M^*(n, p_e)$ by $(M, p_e, \eta)$-code and $M^*(n, p_e, \eta)$, respectively.

For each of the channels considered above, from the channel coding and strong converse theorem due to [9, 10], we know that

$$\lim_{n\to\infty} \frac{1}{n} \log_2 M^*(n, p_e) = C.$$

Thus the first order term in the non-asymptotic expansion of $M^*(n, p_e)$ is $nC$. The second order term in the non-asymptotic expansion of $M^*(n, p_e)$ is given as [6,11]

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

where $V$ is the channel dispersion and $Q^{-1}(.)$ is the inverse of Q-function where Q-function is defined as

$$Q(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-\frac{x^2}{2}} dx.$$

The third order term in the non-asymptotic expansion of $M^*(n, p_e)$, however, depends on the particular channel under discussion (See [12, Theorem 41], [12, Theorem 44], and [12, Theorem 73]). More specifically, for BEC($\epsilon$), we have $C = 1 - \epsilon$, $V = \epsilon(1 - \epsilon)$ and the third order term is $\mathcal{O}(1)$. For BSC($\delta$), we have $C = 1 - h_2(\delta)$, $V = \delta(1 - \delta) \log_2^2(\frac{1-\delta}{\delta})$, and the third order term is $\frac{1}{2} \log n + \mathcal{O}(1)$. For AWGN($\eta$), we have $C = \frac{1}{2} \log_2(1 + \eta)$, $V = \frac{\eta(\eta+2)}{2(\eta+1)^2 \ln^2 2}$, and the third order term is bounded between $\mathcal{O}(1)$ and $\frac{1}{2} \log n + \mathcal{O}(1)$, i.e., the third order term is $\mathcal{O}(\log n)$.

In this paper, we investigate code design over channels with very low capacity. Even though the formula (1) 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 (1), and (ii) some of the terms hidden in $\mathcal{O}(\log n)$ will have significantly higher value and are comparable to the first and second term. In the next section, we will provide non-asymptotic laws for the low-capacity regime.

# 3 Fundamental Limits

**The Low-Capacity Regime.** Consider the transmission over a channel $W$ with capacity $C$. Let $k$ denotes the number of information bits to be sent and $n$ denotes the blocklength of the code. We consider a scenario in which the capacity $C$ is very small, i.e., $C \to 0$. To reliably communicate $k$ bits, we clearly must have $n \geq k/C$ and thus the blocklength $n$ is fairly large. More formally, the low-capacity regime is specified by considering $k$ information bits to be sent over a channel whose capacity $C$ is small and fixed, with blocklength $n$ scaling as $\mathcal{O}(k/C)$ which leads to $k << n$.

In our non-asymptotic derivations we treat the value of $C$ as very small or close to 0 (e.g., $C$ might be less than the probability of error), we treat $n$ as large (i.e., terms such as $1/\sqrt{n}$ are considered as $o(1)$), but the value of $\kappa := nC$ may not be large (e.g., it is in the order of $k$ which can be a few tens). The question that we consider is how does the smallest $\kappa$, for which reliable transmission with error $p_e$ is possible, scale with

$k$? Note that $\kappa$ depends on $n$ through $\kappa = nC$ and finding the smallest (optimal) $\kappa$ is equivalent to finding the smallest length $n$.

A practical situation for low-capacity regime is illustrated next. Consider a wideband AWGN channel with the channel capacity given as $C = B\log(1 + \frac{P}{N_0 B})$ with a fixed total power $P$ and large bandwidth $B$. A wideband user wishes to communicate $k$ bits over this channel in a fixed time frame of duration $T$ seconds. In this scenario, by Nyquist-Shannon sampling theorem, $2BT$ symbols can be transmitted in the given time frame. Suppose that a simple binary phase shift keying (BPSK) modulation is deployed. Hence, the length of transmitted codeword is $n = 2BT$ which is finite but large. As a result, each bit is transmitted through a channel with capacity $\frac{1}{2}\log(1 + \text{SNR}) = \mathcal{O}(\frac{1}{n})$ and the user wishes to transmit $k = \mathcal{O}(1)$ bits over this channel (note that $\text{SNR} = \frac{P}{N_0 B} = \mathcal{O}(\frac{1}{n})$). This implies that the wideband user is operating in an extreme case of the low-capacity regime.

In current low-capacity applications, such as the narrowband and wideband applications discussed in Section 1, the number of information bits $k$ varies between few tens, in narrowband, to few thousands, in wideband, 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 the limit, the low-capacity regime is expressed as follows: We intend to communicate $k$ information bits over a channel with capacity $C \to 0$, and by using a code with length $n \to \infty$. However, the value $\kappa = nC$ stays finite as it will be close to the number of information bits $k$.

**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 (1) are not applicable in the low-capacity regime. Consider transmission over BEC($\epsilon$) with blocklength $n$. When the erasure probability $\epsilon$ 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 later in this section. Another reason for (1) being loose is that some of the terms that are considered as $\mathcal{O}(1)$ will become significant in the low-capacity regime. For instance, we have $1/(\sqrt{n}C) = \sqrt{n}/(nC) = \sqrt{n}/\kappa$ which can not be considered as $o(1)$ since $\kappa$ is usually much smaller than $n$. As we will see, such terms can be captured by using sharper tail inequalities. We will now present and discuss the non-asymptotic laws for channel coding in the low-capacity regime. Proofs of the theorems together with related lemmas are provided in the Appendix.

### 3.1 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. The following theorem provides lower and upper bounds for the best achievable rate in terms of $n$, $p_e$, $\epsilon$, and $\kappa := n(1-\epsilon)$. We use $\mathcal{P}_\lambda(x)$ to denote the Poisson cumulative distribution function, i.e.,

$$\mathcal{P}_\lambda(x) = \Pr\{X < x\}, \qquad \text{where } X \sim \text{Poisson}(\lambda). \tag{2}$$

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

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

*where $M_1$ is the solution of*

$$\mathfrak{P}_1(M_1) + \alpha\sqrt{\mathfrak{P}_1(M_1)} - p_e = 0, \tag{3}$$

*and $M_2$ is the solution of*

$$\mathfrak{P}_2(M_2) - \alpha\sqrt{\mathfrak{P}_2(M_2)} - \alpha\sqrt{\mathcal{P}_\kappa(\log_2 M_2)} - p_e = 0, \tag{4}$$

*and*

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

*Proof.* See Section 5.1 in Appendix. □

The bounds in Theorem 1 are tight and can be computed accurately (see Section 5.2). The bounds 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 $\kappa$ in low-capacity applications varies between a few tens to few hundreds. In such a range, no simple, closed-form approximation of the Poisson distribution with mean $\kappa$ exists. As a result, the lower and upper bounds in Theorem 1 can not be simplified further. Also, 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 4.3 we numerically evaluate the lower and upper bounds predicted by Theorem 1 and compare them with the prediction obtained from Formula (1) [6]. It is observed our predictions are significantly more precise comparing to the prediction obtained from Formula (1) and they become even more precise as the capacity approaches zero.

### 3.2 The Binary Symmetric Channel

Unlike BEC, the non-asymptotic behavior of coding over BSC can be well approximated in low-capacity regime by the central limit theorem (e.g., Berry-Essen theorem). Let us briefly explain why. Consider transmission over BSC($\delta$) where the value of $\delta$ is close to $\frac{1}{2}$. 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 $\delta \to \frac{1}{2}$ one can write $\delta \approx \frac{1}{2} - \sqrt{\frac{\kappa}{n}}$ by using the Taylor expansion of the function $h_2(x)$ around $x = \frac{1}{2}$. Transmission over BSC($\delta$) can be equivalently modeled as follow: (i) With probability $2\delta$ we let the output of the channel be chosen according to Bernoulli($\frac{1}{2}$), 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 $2\delta$ (call it the noisy event) and completely noiseless with probability $1 - 2\delta$ (call it the noiseless event). As $\delta \to \frac{1}{2}$, then the noiseless even 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 (1) 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 Low-Capacity BSC)**.** *Consider transmission over BSC($\delta$) in low-capacity regime and let $\kappa = n(1 - h_2(\delta))$. Then,*

$$\log_2 M^*(n, p_e) = \kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}} \cdot \sqrt{\kappa}\, Q^{-1}(p_e) + \frac{1}{2}\log_2 \kappa - \log_2 p_e + \mathcal{O}(\log\log\kappa)\,. \tag{5}$$

*Proof.* See Section 5.3 in Appendix. □

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.[3] We also note that, similar to the BEC case, all terms in (5) are expressed in terms of $\kappa$ 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 $n$.

**Corollary 1.** *Consider transmission of $k$ information bits over a low-capacity BSC($\delta$). Then, the optimal blocklength $n^*$ for such a transmission is*

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

*Proof.* See Section 5.3 in Appendix. □

### 3.3 The Additive White Gaussian Channel

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 [12, Theorem 73] with slight modifications.

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

$$\log_2 M^*(n, p_e, \eta) = \kappa - \frac{\sqrt{\eta+2}}{(\eta+1)\sqrt{\ln 2}} \cdot \sqrt{\kappa}\, Q^{-1}(p_e) + \mathcal{E}, \tag{6}$$

*where*

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

*Proof.* See Section 5.4 in Appendix. □

Same considerations about $\mathcal{O}(.)$ notation as discussed earlier, should be taken into account here. Also note that as for BEC and BSC, the optimal blocklength for 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($\eta$). Then, the optimal blocklength $n^*$ for such a transmission is*

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

*Proof.* See Section 5.4 in Appendix. □

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

# 4 Practical Code Designs and Simulation Results

As we need to design codes with extremely low rate, some of the stat-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. E.g., an $(l, r)$-regular LDPC code with design rate $R = 0.01$ requires $r, l \geq 99$. Hence, the Tanner graph will have minimum degree of at least 99 and even for codelengths of order tens of thousands the Tanner graph will have many short cycles. In order to circumvent this issue, the current practical designs, e.g., the NB-IoT code design, use repetition coding. I.e., 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 4.1, we will discuss the pros and cons of using repetition schemes along with trade-offs between the number of repetitions and performance of the code. As we will see, although repetition leads to efficient implementations, the rate loss through many repetitions will result in codes with mediocre performance.

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 4.2, we will study the behaviour of polar coding on low-capacity channels. As we will see, polar coding is advantageous in terms of distance, performance and implicit repetition, however, its encoding and decoding algorithms have to be carefully adjusted to reduce complexity and latency for practical applications.

Throughout this section, we will consider code design for the class of binary memoryless symmetric (BMS) channels. A BMS channel $W$ has binary input and, letting $W(y \mid x)$ denotes the transition matrix, there exists a permutation $\pi$ 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.

## 4.1 How Much Repetition is Needed?

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 which is the result of passing $r$ copies of the input bit independently through the original channel $W$. E.g., if $W$ is BEC($\epsilon$) 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 (especially if $r$ 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 the encoding side, we just need to repeat each of its code 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 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 way.

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 any more. 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 is the largest repetition size $r_\beta$ such that

$$\frac{n}{r_\beta} C(W^{r_\beta}) \geq \beta n C(W). \tag{7}$$

Let us first assume that transmission takes place over BEC($\epsilon$). We thus have $W^r = \text{BEC}(\epsilon^r)$. If $\epsilon$ is not close to 1, then even $r = 2$ would result in a considerble rate loss, e.g., if $\epsilon = 0.5$, then $C(W^2) = 0.75$ whereas $2C(W) = 1$. However, when $\epsilon$ 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, $r_\beta$, that satisfies (7).

**Theorem 4** (Maximum Repetition Length for BEC)**.** *If $W = BEC(\epsilon)$, then for the largest repetition size, $r_\beta$, that satisfies* (7)*, we have*

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

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

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

*Proof.* See Section 5.5 in Appendix. □

**Remark 1.** *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 4 we can conclude that, when $\epsilon$ 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 $\kappa$ 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 loss. We refer to Section 4.3 for a numerical case study on repetition coding. 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.

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 $r_\beta$ 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 $r_\beta$ that satisfies* (7)*. Hence, for any BMS channel with capacity C and $\kappa = nC$, we have*

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

*Proof.* See Section 5.5 in Appendix. □

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

### 4.2 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 will describe such advantages together with supporting analytical and numerical evidence. We also show later in this section that, in order to make polar codes a suitable candidate for practice, we need to carefully adapt their encoding and decoding operations. We begin by providing a brief description of polar codes to set up notation and the basics.

**Basics of Polar Coding [7].** The basis of channel polarization consists in 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

$$W^0(y_1, y_2 \mid x_1) = \sum_{x_2 \in \mathcal{X}} \frac{1}{2} W(y_1 \mid x_1 \oplus x_2) W(y_2 \mid x_2), \tag{9}$$

$$W^1(y_1, y_2, x_1 \mid x_2) = \frac{1}{2} W(y_1 \mid x_1 \oplus x_2) W(y_2 \mid x_2). \tag{10}$$

Then, $W^0$ is a worse channel in the sense that it is degraded with respect to $W$, hence less reliable than $W$; and $W^1$ is a better channel in the sense that it is upgraded with respect to $W$, hence more reliable than $W$. In the polar coding literature, the operation in (9) is also known as the *check* or *minus* operation and the operation in (10) 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 $\{W_m^{(i)}\}_{i \in \{0,\ldots,n1\}}$. More specifically, given $i \in \{0, \ldots, n-1\}$, let $(b_1, b_2, \ldots, b_m)$ be its binary expansion over $m$ bits, where $b_1$ is the most significant bit and $b_m$ is the least significant bit, i.e.,

$$i = \sum_{k=1}^{m} b_k 2^{m-1-k}.$$

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

$$W_n^{(i)} = (((W^{b_1})^{b_2})^{\cdots})^{b_m}.$$

**Example 1** (Synthetic Channel)**.** *Take* $m = 4$ *and* $i = 10$*. Then, the synthetic channel* $W_{16}^{(10)} = (((W^1)^0)^1)^0$ *is obtained by applying first* (10)*, then* (9)*, then* (10)*, and finally* (9).

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

$$G_n = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes m},$$

which correspond to the best $k$ synthetic channels. It is worth noting that for an index $i$ with binary expansion $(b_1, b_2, \cdots, b_n)$ the Hamming wight of the $i$-th row of $G_n$ is $2^{\sum_{j=1}^{m} b_i}$, i.e., the Hamming weight of the $i$-th row, which corresponds to the $i$-th synthetic channel, is exponentially related with number of plus operations in the construction of the $i$-th synthetic channel.

**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

| $n$ | 1024 | 2048 | 4096 | 8192 | 16364 |
|---|---|---|---|---|---|
| $d_{\min}(d_{\min}/n)$ | 128 (1/8) | 256 (1/8) | 512 (1/8) | 1024 (1/8) | 2048 (1/8) |

Table 1: Minimum distance of a polar code constructed for $k = 40$ over various channels with capacity 0.02.

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 1 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 $k = 40$. For the range of $n$ shown in the table, we have observed that the set of synthetic indices for all the three channels were identical. This would suggest the *universality* of polar codes in the low-rate regime (this should only hold when $k << n$).[4] As the table shows, the minimum distance keeps increasing linearly with $n$.

**Polar Coding Does Optimal and Implicit Repetition at Low-Capacity.** We have shown in Section 4.1 that the maximum allowed repetition size to have 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, the simplified SC decoder [16] or relaxed polar codes [17] can be invoked. Such complexity reductions are important for making polar codes a suitable candidate for practice.

**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 $m_0$ 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 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.*

### 4.3 Simulation Results

For the BEC, we have compared in Figure **1**, the lower and upper bounds obtained from Theorem 1 with the predictions of Formula (1). 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($\epsilon$). 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 e.g., 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 4 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. This is natural because the upper bound has been obtained by a random coding achievability argument. As expected, the prediction obtained from Formula (1) is not precise in the low-capacity regime and it becomes worse as the capacity approaches

[4]Note that polar codes are not universal in general [13], but universal polar-like constructions exist [14, 15].

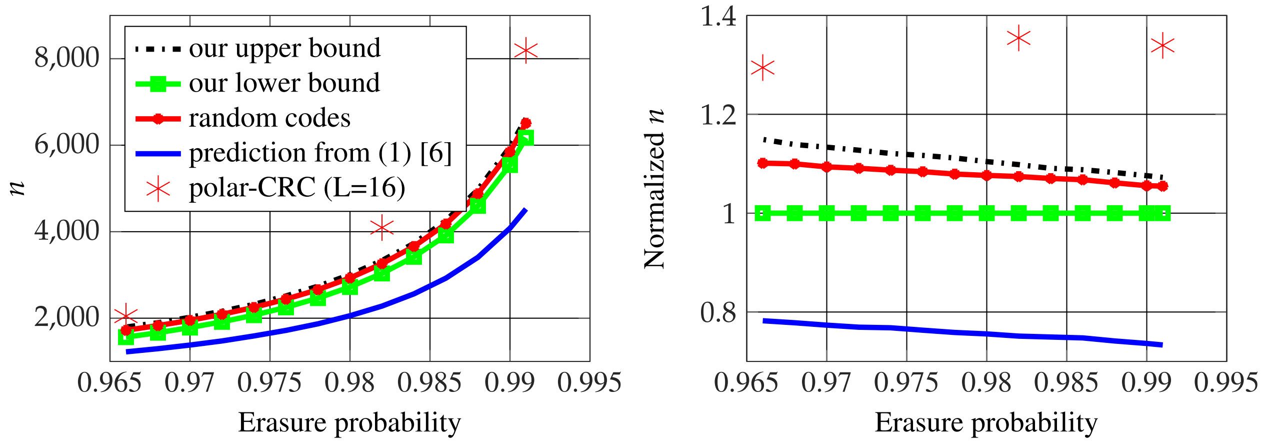


**Figure 1**: Comparison for low-capacity BEC. The number of information bits is $k = 40$ and the target error probability is $p_e = 10^{-2}$. For the right plot, with the same legend entries as the left plot, all the blocklengths $n$ in the left plot are normalized by the value of the lower bound, obtained from Theorem 1.

zero. Also, the performance of 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 [18] 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 $p_e = 10^{-2}$. We have compared in Figure **2**, the predictions from Theorem 2 and Formula (1). As we expected, the prediction from Formula (1) is quite imprecise in the low-capacity regime. Note that the prediction of Theorem 2 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 analyzie the finite-length scaling of polar codes in the low-capacity regime [19–23].

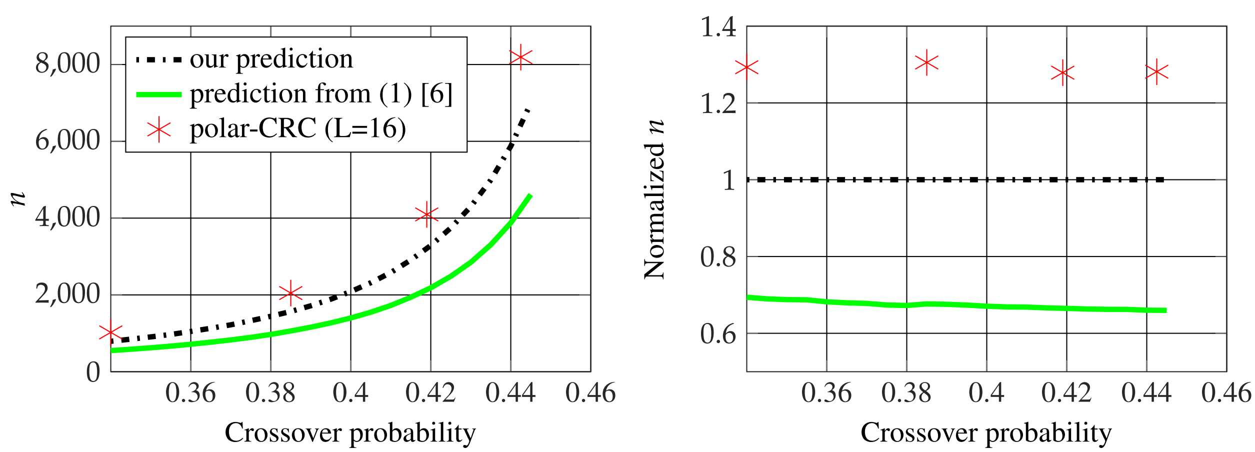


**Figure 2**: Comparison for low-capacity BSC. The number of information bits is $k = 40$ and the target error probability is $p_e = 10^{-2}$. For the right plot, with the same legend entries as the left plot, all the blocklengths $n$ in the left plot are normalized by the value of the prediction obtained from Theorem 2.

Figure **3** compares the performance of polar codes with repeated LTE Turbo codes over the binary-input additive white Gaussian 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 6. Hence, the polar coding scheme is actually a $(2048, 40)$ polar code with 4 repetitions. We note from Section 1 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 **3**. As we see from the figure, the waterfall region of Turbo-repitition 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 a code carefully designed, both in terms of construction as well as the number of repetitions, for the total length 8192, such as the considered polar code.

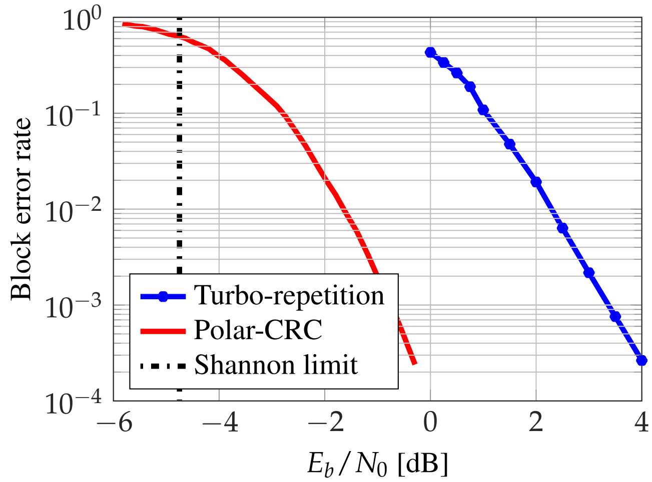


**Figure 3**: Comparison for low-capacity BAWGN. The number of information bits is $k = 40$. The polar-CRC has length 8192, is constructed using 6 CRC bits, and is decoded using the SC-list decoder with $L = 16$. The Turbo-repitition has an underlying $(120, 40)$ Turbo code which is repeated 68 times (total length $= 8160$) and is decoded with 6 iterations. The Shannon limit for this setting is $-4.75$ dB.

# 5 Appendix: Proofs

## 5.1 Proofs for BEC

In this section we will prove the converse and achievability bounds of Theorem 1. In the proofs we will be using Theorems 7–10 which are stated at the end of this section. For results in coding theory, we generally refer to [12] as it has well collected and presented the corresponding proofs. See also [8], [24], [25], [26], and [27].

**Proof of Theorem 1. Achievability Bound.** Consider $n$ transmissions over the BEC($\epsilon$) which are indexed by $i = 1, \cdots, n$. For the $i$-th transmission, we let $X_i$ be a Bernoulli random variable which is 0 if the output of the $i^{\text{th}}$ channel is an erasure and is 1 otherwise, i.e., $\Pr\{X_i = 1\} = 1 - \epsilon$. Suppose $S_n = \sum_{i=1}^{n} X_i$ and denote $\kappa = n(1 - \epsilon)$. We will use the result of Theorem 10 and show that if a number $M_1$ satisfies (3), then it will dissatisfy the inequality in (28). As a result, we obtain $M_1 \le M^*(n, p_e)$. Now, by considering (28), we define

$$I_1 = \sum_{r < \log_2 M_1} \binom{n}{r} \epsilon^{n-r}(1-\epsilon)^r,$$
$$I_2 = \sum_{\log_2 M_1 \le r \le n} \binom{n}{r} \epsilon^{n-r}(1-\epsilon)^r \frac{M_1}{2^r}.$$

We have $I_1 = \Pr\{S_n < \log_2 M_1\}$. Suppose $X \sim \text{Poisson}(\kappa)$, then we can write

$$\begin{aligned} I_1 &\le \Pr\{X < \log_2 M_1\} + \Big|\Pr\{S_n < \log_2 M_1\} - \Pr\{X < \log_2 M_1\}\Big|, \\ I_2 &= \sum_{\log_2 M_1 \le r \le n} \Pr\{S_n = r\} \frac{M_1}{2^r} \\ &\le \sum_{\log_2 M_1 \le r \le n} \Pr\{X = r\} \frac{M_1}{2^r} + \sum_{\log_2 M_1 \le r \le n} \Big|\Pr\{X = r\} - \Pr\{S_n = r\}\Big| \frac{M_1}{2^r} \\ &\le \sum_{\log_2 M_1 \le r < \infty} \Pr\{X = r\} \frac{M_1}{2^r} + \sum_{\log_2 M_1 \le r < \infty} \Big|\Pr\{X = r\} - \Pr\{S_n = r\}\Big| \frac{M_1}{2^r}. \end{aligned}$$

Putting these together, we obtain

$$I_1 + I_2 \le \mathbb{E}\left[\mathbf{1}(X < \log_2 M_1) + \mathbf{1}(X \ge \log_2 M_1) \frac{M_1}{2^X}\right] + J_1 + J_2, \tag{11}$$

where

$$J_1 = \Big|\Pr\{S_n < \log_2 M_1\} - \Pr\{X < \log_2 M_1\}\Big|,$$
$$J_2 = \sum_{\log_2 M_1 \le r < \infty} \Big|\Pr\{X = r\} - \Pr\{S_n = r\}\Big| \frac{M_1}{2^r}.$$

Using Theorem 7, we have

$$J_1 = \Big|\Pr\{S_n < \log_2 M_1\} - \Pr\{X < \log_2 M_1\}\Big| \le \alpha_1 \sqrt{\Pr\{X < \log_2 M_1\}}, \tag{12}$$
$$\Big|\Pr\{X = r\} - \Pr\{S_n = r\}\Big| \le \alpha_2 \sqrt{\Pr\{X \le r\}}, \tag{13}$$

where

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

We now upper-bound $J_2$ using (13), as follows:

$$\begin{aligned} J_2 &= 2 \sum_{\log_2 M_1 \le r < \infty} \frac{M_1}{2^{r+1}} \Big| \Pr\{X = r\} - \Pr\{S_n = r\} \Big| \\ &\le 2\alpha_2 \sum_{\log_2 M_1 \le r < \infty} \frac{M_1}{2^{r+1}} \sqrt{\Pr\{X \le r\}} \\ &\le 2\alpha_2 \sqrt{\sum_{\log_2 M_1 \le r < \infty} \frac{M_1}{2^{r+1}} \Pr\{X \le r\}}. \end{aligned} \tag{14}$$

For obtaining the last part, note that $\sqrt{x}$ is a concave function and

$$\sum_{\log_2 M_1 \le r < \infty} \frac{M_1}{2^{r+1}} = 1.$$

Thus, (14) follows from Jensen inequality for $\sqrt{x}$, that is, $\mathbb{E}\left[\sqrt{Z}\right] \le \sqrt{\mathbb{E}[Z]}$. Also consider

$$\begin{aligned} \sum_{\log_2 M_1 \le r < \infty} \frac{M_1}{2^{r+1}} \Pr\{X \le r\} &= \sum_{\log_2 M_1 \le r < \infty} \frac{M_1}{2^{r+1}} \sum_{i=0}^{r} e^{-\kappa} \frac{\kappa^i}{i!} \\ &= \sum_{i=0}^{\infty} e^{-\kappa} \frac{\kappa^i}{i!} \left( \mathbf{1}(i < \log_2 M_1) + \frac{\mathbf{1}(i \ge \log_2 M_1)}{2^{i - \log_2 M_1}} \right) \\ &= \sum_{i < \log_2 M_1} e^{-\kappa} \frac{\kappa^i}{i!} + M_1 \sum_{i \ge \log_2 M_1} e^{-\kappa} \frac{\kappa^i}{i!} \frac{1}{2^i} \\ &= \mathcal{P}_\kappa(\log_2 M_1) + M_1 e^{-\kappa/2} \sum_{i \ge \log_2 M_1} e^{-\kappa/2} \frac{(\kappa/2)^i}{i!} \\ &= \mathcal{P}_\kappa(\log_2 M_1) + M_1 e^{-\kappa/2} \left(1 - \mathcal{P}_{\kappa/2}(\log_2 M_1)\right). \end{aligned} \tag{15}$$

From (14) and (15), we arrive at

$$J_2 \le 2\alpha_2 \sqrt{\mathcal{P}_\kappa(\log_2 M_1) + M_1 e^{-\kappa/2} \left(1 - \mathcal{P}_{\kappa/2}(\log_2 M_1)\right)}. \tag{16}$$

Also, considering the notation in (2), we can write

$$\begin{aligned} &\mathbb{E}\left[ \mathbf{1}(X < \log_2 M_1) + \mathbf{1}(X \ge \log_2 M_1) \frac{M_1}{2^X} \right] \\ &\quad = \mathbb{E}\Big[ \mathbf{1}(X < \log_2 M_1) \Big] + \mathbb{E}\left[ \frac{M_1}{2^X} \right] - \mathbb{E}\left[ (X < \log_2 M_1) \frac{M_1}{2^X} \right] \\ &\quad = \mathcal{P}_\kappa(\log_2 M_1) + \sum_r e^{-\kappa} \frac{\kappa^r}{r!} \cdot \frac{M_1}{2^r} - \sum_{r < \log_2 M_1} e^{-\kappa} \frac{\kappa^r}{r!} \cdot \frac{M_1}{2^r} \\ &\quad = \mathcal{P}_\kappa(\log_2 M_1) + M_1 e^{-\kappa/2} \sum_r e^{-\kappa/2} \frac{(\kappa/2)^r}{r!} - M_1 e^{-\kappa/2} \sum_{r < \log_2 M_1} e^{-\kappa/2} \frac{(\kappa/2)^r}{r!} \\ &\quad = \mathcal{P}_\kappa(\log_2 M_1) + M_1 e^{-\kappa/2} - M_1 e^{-\kappa/2} \mathcal{P}_{\kappa/2}(\log_2 M_1). \end{aligned} \tag{17}$$

Now, (11), (12), (16), and (17) together result in

$$\begin{aligned} I_1 + I_2 &\le \mathcal{P}_\kappa(\log_2 M_1) + M_1 e^{-\kappa/2}\left(1 - \mathcal{P}_{\kappa/2}(\log_2 M_1)\right) \\ &\quad + \alpha_1 \sqrt{\mathcal{P}_\kappa(\log_2 M_1)} + 2\alpha_2 \sqrt{\mathcal{P}_\kappa(\log_2 M_1) + M_1 e^{-\kappa/2}\left(1 - \mathcal{P}_{\kappa/2}(\log_2 M_1)\right)} \\ &\le \mathfrak{P}_1(M_1) + \alpha \sqrt{\mathfrak{P}_1(M_1)} \\ &= p_e, \end{aligned} \tag{18}$$

where

$$\begin{aligned} \mathfrak{P}_1(M_1) &= \mathcal{P}_\kappa(\log_2 M_1) + M_1 e^{-\kappa/2}\left(1 - \mathcal{P}_{\kappa/2}(\log_2 M_1)\right), \\ \alpha = \alpha_1 + 2\alpha_2 &= \frac{\sqrt{2}}{\epsilon^{3/2}}\left(1 + 2\sqrt{\frac{3}{\epsilon\kappa}}\right)\left(\sqrt{e} - 1\right)(1 - \epsilon). \end{aligned}$$

Note that (18) holds by the definition of $M_1$ in (3). Hence, we showed

$$\sum_{r=0}^{n} \binom{n}{r} \epsilon^r (1-\epsilon)^{n-r} 2^{-[r - \log_2(M_1 - 1)]^+} \le I_1 + I_2 \le p_e,$$

which means $M_1$ dissatisfies the inequality in (28). Hence, $M_1 \le M^*(n, p_e)$. □

**Proof of Theorem 1. Converse Bound.** Consider $n$ transmissions over the BEC($\epsilon$) which are indexed by $i = 1, \cdots, n$. For the $i$-th transmission, we let $X_i$ be a Bernoulli random variable which is 0 if the output of the $i^{\text{th}}$ channel is an erasure and is 1 otherwise, i.e., $\Pr\{X_i = 1\} = 1 - \epsilon$. Suppose $S_n = \sum_{i=1}^n X_i$ and denote $\kappa = n(1-\epsilon)$. We will use the result of Theorem 9 and show that if a number $M_2$ satisfies (4), then it will dissatisfy the inequality in (27). As a result, we obtain $M^*(n, p_e) \le M_2$. Now, by considering (27), we define

$$\begin{aligned} I_1 &= \sum_{r < \log_2 M_2} \binom{n}{r} \epsilon^{n-r} (1-\epsilon)^r, \\ I_2 &= \sum_{r < \log_2 M_2} \binom{n}{r} \epsilon^{n-r} (1-\epsilon)^r \frac{2^r}{M_2}. \end{aligned}$$

We have $I_1 = \Pr\{S_n < \log_2 M_2\}$. Suppose $X \sim \text{Poisson}(\kappa)$, then we can write

$$\begin{aligned} I_1 &\ge \Pr\{X < \log_2 M_2\} - \left| \Pr\{S_n < \log_2 M_2\} - \Pr\{X < \log_2 M_2\} \right|, \\ I_2 &= \sum_{r < \log_2 M_2} \Pr\{S_n = r\} \frac{2^r}{M_2} \\ &\le \sum_{r < \log_2 M_2} \Pr\{X = r\} \frac{2^r}{M_2} + \sum_{r < \log_2 M_2} \left| \Pr\{X = r\} - \Pr\{S_n = r\} \right| \frac{2^r}{M_2}. \end{aligned}$$

Putting these together, we obtain

$$I_1 - I_2 \ge \mathbb{E}\left[ \mathbf{1}\left(X < \log_2 M_2\right) \left(1 - \frac{2^X}{M_2}\right) \right] - J_1 - J_2, \tag{19}$$

where

$$J_1 = \left| \Pr\left\{ S_n < \log_2 M_2 \right\} - \Pr\left\{ X < \log_2 M_2 \right\} \right|,$$

$$J_2 = \sum_{r < \log M_2} \left| \Pr\left\{ X = r \right\} - \Pr\left\{ S_n = r \right\} \right| \frac{2^r}{M_2}.$$

Using Theorem 7, we have

$$J_1 = \left| \Pr\left\{ S_n < \log_2 M_2 \right\} - \Pr\left\{ X < \log_2 M_2 \right\} \right| \le \alpha_1 \sqrt{\Pr\left\{ X < \log_2 M_2 \right\}} \tag{20}$$

$$\left| \Pr\left\{ X = r \right\} - \Pr\left\{ S_n = r \right\} \right| \le \alpha_2 \sqrt{\Pr\left\{ X \le r \right\}}, \tag{21}$$

where

$$\alpha_1 = \frac{\sqrt{2}}{\epsilon^{3/2}} \left( \sqrt{e} - 1 \right) (1 - \epsilon),$$

$$\alpha_2 = \frac{\sqrt{6}}{\epsilon^2 \sqrt{\kappa}} \left( \sqrt{e} - 1 \right) (1 - \epsilon).$$

We now upper-bound $J_2$ using (21), as follows:

$$\begin{aligned} J_2 &= \sum_{r < \log M_2} \frac{2^r}{M_2} \left| \Pr\left\{ X = r \right\} - \Pr\left\{ S_n = r \right\} \right| \\ &\le \alpha_2 \sum_{r < \log M_2} \frac{2^r}{M_2} \sqrt{\Pr\left\{ X \le r \right\}} \\ &\le \alpha_2 \sqrt{\sum_{r < \log M_2} \frac{2^r}{M_2} \Pr\left\{ X \le r \right\}}. \end{aligned} \tag{22}$$

For obtaining the last part, note that $\sqrt{x}$ is a concave function and

$$\sum_{r < \log M_2} \frac{2^r}{M_1} \le 1.$$

Thus, (22) follows from Jensen inequality for $\sqrt{x}$, that is, $\mathbb{E}\left[ \sqrt{Z} \right] \le \sqrt{\mathbb{E}\left[ Z \right]}$. Also consider

$$\begin{aligned} \sum_{r < \log M_2} \frac{2^r}{M_2} \Pr\left\{ X \le r \right\} &= \sum_{r < \log M_2} \frac{2^r}{M_2} \left( 1 - \Pr\left\{ X > r \right\} \right) \\ &= 1 - \frac{1}{M_2} - \sum_{r < \log M_2} \frac{2^r}{M_2} \sum_{i=r+1}^{\infty} e^{-\kappa} \frac{\kappa^i}{i!} \\ &= 1 - \frac{1}{M_2} - \sum_{i=0}^{\infty} e^{-\kappa} \frac{\kappa^i}{i!} \left( \mathbf{1}\left( i \ge \log_2 M_2 \right) + \mathbf{1}\left( i < \log_2 M_2 \right) \frac{2^i - 1}{M_2} \right) \\ &= 1 - \frac{1}{M_2} - \left( 1 - \mathcal{P}_\kappa(\log_2 M_2) + \frac{e^\kappa}{M_2} \sum_{i < \log M_2} e^{-2\kappa} \frac{(2\kappa)^i}{i!} - \frac{\mathcal{P}_\kappa(\log_2 M_2)}{M_2} \right) \\ &= \mathcal{P}_\kappa(\log_2 M_2) - \frac{e^\kappa}{M_2} \mathcal{P}_{2\kappa}\left( \log_2 M_2 \right) - \frac{1 - \mathcal{P}_\kappa(\log_2 M_2)}{M_2} \\ &\le \mathcal{P}_\kappa(\log_2 M_2) - \frac{e^\kappa}{M_2} \mathcal{P}_{2\kappa}\left( \log_2 M_2 \right). \end{aligned} \tag{23}$$

From (22) and (23), we arrive at

$$J_2 \le \alpha_2 \sqrt{\mathcal{P}_\kappa(\log_2 M_2) - \frac{e^\kappa}{M_2}\, \mathcal{P}_{2\kappa}\left(\log_2 M_2\right)}. \tag{24}$$

Also, considering the notation in (2), we can write

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}\left(X < \log_2 M_2\right)\left(1 - \frac{2^X}{M_2}\right)\right] &= \mathbb{E}\left[\mathbf{1}\left(X < \log_2 M_2\right)\right] - \mathbb{E}\left[\left(X < \log_2 M_2\right)\frac{2^X}{M_2}\right] \\
&= \mathcal{P}_\kappa(\log_2 M_2) - \sum_{r < \log_2 M_2} e^{-\kappa}\frac{\kappa^r}{r!}\cdot\frac{2^r}{M_2} \\
&= \mathcal{P}_\kappa(\log_2 M_2) - \frac{e^\kappa}{M_2}\sum_{r < \log_2 M_2} e^{-2\kappa}\frac{(2\kappa)^r}{r!} \\
&= \mathcal{P}_\kappa(\log_2 M_2) - \frac{e^\kappa}{M_2}\,\mathcal{P}_{2\kappa}\left(\log_2 M_2\right). \qquad (25)
\end{aligned}$$

Now, (19), (20), (24), and (25) together result in

$$\begin{aligned}
I_1 - I_2 &\ge \mathcal{P}_\kappa(\log_2 M_2) - \frac{e^\kappa}{M_2}\,\mathcal{P}_{2\kappa}\left(\log_2 M_2\right) - \alpha_1\sqrt{\mathcal{P}_\kappa(\log_2 M_2)} - \alpha_2\sqrt{\mathcal{P}_\kappa(\log_2 M_2) - \frac{e^\kappa}{M_2}\,\mathcal{P}_{2\kappa}\left(\log_2 M_2\right)} \\
&\ge \mathfrak{P}_2(M_2) - \alpha\sqrt{\mathfrak{P}_2(M_2)} - \alpha\sqrt{\mathcal{P}_\kappa(\log_2 M_2)} \\
&= p_e, \qquad (26)
\end{aligned}$$

where

$$\begin{aligned}
\mathfrak{P}_2(M_2) &= \mathcal{P}_\kappa(\log_2 M_2) - \frac{e^\kappa}{M_2}\,\mathcal{P}_{2\kappa}\left(\log_2 M_2\right), \\
\alpha = \alpha_1 + 2\alpha_2 &= \frac{\sqrt{2}}{\epsilon^{3/2}}\left(1 + 2\sqrt{\frac{3}{\epsilon\kappa}}\right)\left(\sqrt{e} - 1\right)(1 - \epsilon).
\end{aligned}$$

Note that (26) holds by the definition of $M_2$ in (4). Hence, we showed

$$\sum_{r < \log_2 M_2} \binom{n}{r} \epsilon^{n-r}(1-\epsilon)^r \left(1 - \frac{2^r}{M_2}\right) = I_1 - I_2 \ge p_e,$$

which means $M_2$ dissatisfies the inequality in (27). Hence, $M^*(n, p_e) \le M_2$. □

**Theorem 7** (Strong Poisson Convergence)**.** *For $1 \le i \le n$ let $X_i$ be independent random variables with $P(X_i = 1) = 1 - P(X_i = 0) = p_i$. Define $S_n = \sum_{i=1}^n X_i$. Also define $\lambda_k = \sum_{i=1}^n p_i^k$ and let $\lambda = \lambda_1$ and $\theta = \lambda_2/\lambda_1$. Then,*

$$\left| Pr\{S_n \le m\} - Pr\{X \le m\} \right| \le \frac{\sqrt{2}\left(\sqrt{e} - 1\right)\theta}{(1-\theta)^{3/2}}\sqrt{\psi(m)},$$

*and*

$$\left| Pr\{S_n = m\} - Pr\{X = m\} \right| \le \frac{\sqrt{6}\left(\sqrt{e} - 1\right)\theta}{(1-\theta)^2\sqrt{\lambda}}\sqrt{\psi(m)},$$

*where $X \sim$ Poisson$(\lambda)$ and the quantity $\psi(m)$ is*

$$\psi(m) = \min\left\{Pr\{X \le m\}, Pr\{X > m\}\right\}.$$

*Proof.* See [28, Theorem 3.4, Lemma 3.7]. □

**Theorem 8** (Poisson Convergence)**.** *For $1 \leq i \leq n$ let $X_i$ be independent random variables with $P(X_i = 1) = 1 - P(X_i = 0) = p_i$. Define $S_n = \sum_{i=1}^n X_i$ and $\lambda_k = \sum_{i=1}^n p_i^k$. Let $\mu_n$ be the distribution of $S_n$ and $\nu_n$ be the Poisson distribution with mean $\lambda_1$. Then the following holds.*

$$\sup_A \left| \mu_n(A) - \nu_n(A) \right| \leq \frac{\lambda_2}{\lambda_1}$$

*Proof.* See [29, page 89]. □

**Theorem 9** (Converse Bound for BEC)**.** *For any $(M, p_e)$-code over the BEC$^n(\epsilon)$, we have*

$$p_e \geq \sum_{r < \log_2 M} \binom{n}{r} \epsilon^{n-r} (1-\epsilon)^r \left(1 - \frac{2^r}{M}\right). \tag{27}$$

*Proof.* See [12, Theorem 43]. □

**Theorem 10** (RCU Achievability Bound for BEC)**.** *There exists an $(M, p_e)$-code over BEC$^n(\epsilon)$ such that*

$$p_e \leq \sum_{r=0}^{n} \binom{n}{r} \epsilon^r (1-\epsilon)^{n-r} 2^{-[r - \log_2(M-1)]^+}. \tag{28}$$

*Proof.* See [12, Corollary 42]. □

## 5.2 How to Compute the Bounds in Theorem 1

The problem essentially boils down to accurate computation of the probabilities $\Pr(X = s)$ when $X$ is a Poisson random variable with average $\kappa$. We have $\Pr(X = s) = e^{-\kappa}\kappa^s / s!$. The value of $s!$ can be approximated using the refined Ramanujan's formula [30]:

$$s! = \sqrt{\pi} \left(\frac{s}{e}\right)^s \left(8s^3 + 4s^2 + s + \frac{\theta(s)}{30}\right)^{\frac{1}{6}},$$

where

$$\theta_1(s) := 1 - \frac{11}{8s} + \frac{79}{112s^2} \leq \theta(s) \leq \theta_2(s) := 1 - \frac{11}{8s} + \frac{79}{112s^2} + \frac{20}{33s^3}.$$

By plugging-in the lower bound for $s!$ from the above formula (for $s \leq 10$ we can use the exact value of $s!$) we obtain

$$\Pr(X = s) = \frac{e^{s - \kappa + s\ln(\kappa/s)}}{\sqrt{\pi}\left(8s^3 + 4s^2 + s + \frac{\theta_1(s)}{30}\right)^{\frac{1}{6}}}.$$

Note that this formula is exact up to a multiplicative factor of $1 + s^{-6}$ which for $s \geq 10$ gives us a $(1 + 10^{-6})$-approximation. Moreover, if we are obsessed with obtaining "bounds", we can use the lower and upper bound approximations for $s!$ to bound the Poisson probability from above and below and hence obtain bounds for $M_1, M_2$ in Theorem 1.

Once the Poisson probabilities are approximated (or bounded) suitably, we can compute values of $M_1$ and $M_2$ up to any precision by truncating the expectation and using the bisection method to solve the equations.

### 5.3 Proofs for BSC

In this section we will prove the converse and achievability bounds of Theorem 2. In the proofs we will be using Theorems 11–14 as well as Lemmas 1–4 which are stated at the end of this section. For results in coding theory, we generally refer to [12] as it has well collected and presented the corresponding proofs. See also [8], [24], [25], [26], and [27].

**Proof of Theorem 2. Achievability Bound.** Define $T$ and $S$ as follows:

$$T = n\delta + \sqrt{n\delta(1-\delta)}Q^{-1}\left(p_e - \frac{p_e}{\sqrt{\kappa}} - \frac{c_1}{\sqrt{n}}\right), \tag{29}$$

$$S = \sum_{r \le T} \binom{n}{r} \delta^r (1-\delta)^{n-r} S_n^r, \tag{30}$$

where $c_1 = \frac{3(2\delta^2 - 2\delta + 1)}{\sqrt{\delta(1-\delta)}}$ is a constant. Suppose we choose some $M$ to satisfy

$$M \le \frac{p_e}{\sqrt{\kappa}\, S}. \tag{31}$$

Now define

$$I_1 = \sum_{r \le T} \binom{n}{r} \delta^r (1-\delta)^{n-r} M S_n^r,$$

$$I_2 = \sum_{T < r \le n} \binom{n}{r} \delta^r (1-\delta)^{n-r}.$$

From (30) and (31), we have

$$I_1 = MS \le \frac{p_e}{\sqrt{\kappa}}. \tag{32}$$

For $1 \le i \le n$, deine $V_i$ to be Bernoulli random variables with $\Pr\{V_i = 1\} = 1 - \Pr\{V_i = 0\} = \delta$. Then define

$$\mu = \mathbb{E}\left[V_i\right] = \delta, \quad \sigma^2 = \mathbb{V}ar[V_i] = \delta(1-\delta), \quad \rho = \mathbb{E}\left[|V_i - \mu|^3\right] = \delta(1-\delta)\left(2\delta^2 - 2\delta + 1\right).$$

Now, using Theorem 11, we can write

$$I_2 = \Pr\left\{\sum_{i=1}^{n} V_i > T\right\} \le Q\left(\frac{T - n\mu}{\sigma}\right) + \frac{3\rho}{\sigma^3 \sqrt{n}} = Q\left(\frac{T - n\delta}{\sqrt{n\delta(1-\delta)}}\right) + \frac{c_1}{\sqrt{n}} \stackrel{(29)}{=} p_e - \frac{p_e}{\sqrt{\kappa}}. \tag{33}$$

Therefore, (32) and (33) together give

$$I_1 + I_2 \le p_e - \frac{p_e}{\sqrt{\kappa}} + \frac{p_e}{\sqrt{\kappa}} = p_e.$$

Now, we can write

$$\sum_{r=0}^{n} \binom{n}{r} \delta^r (1-\delta)^{n-r} \min\left\{1, (M-1)S_n^r\right\} \le \sum_{r=0}^{n} \binom{n}{r} \delta^r (1-\delta)^{n-r} \min\left\{1, M S_n^r\right\} \le I_1 + I_2 \le p_e.$$

Thus, $M$ dissatisfies the inequality in (48) and as a result, $M^*(n, p_e) \geq M$. Note that $M$ was arbitrarily chosen to satisfy (31), This means for any $M$ satisfying (31), we have $M^*(n, p_e) \geq M$. Hence,

$$M^*(n, p_e) \geq \sup \left\{ M : M \leq \frac{p_e}{\sqrt{\kappa}\, S} \right\} = \frac{p_e}{\sqrt{\kappa}\, S}. \tag{34}$$

Now, in order to find a lower bound for $M^*(n, p_e)$, it suffices to find an upper bound for $S$. This is our main goal for the rest of the proof. Note that due to Lemma 3, there exists a constant $\theta$ such that

$$\binom{n}{r} \delta^r (1-\delta)^{n-r} \leq \frac{\theta}{\sqrt{n}}. \tag{35}$$

Also note that $S_n^r$ is increasing with respect to $r$. Define

$$\beta = \frac{1}{4} \sqrt{\frac{\ln 2}{2}} \cdot \frac{\log_2 \kappa}{\sqrt{\kappa}}.$$

With this choice of $\beta$, we continue as follows:

$$\begin{aligned} S &= \sum_{r \leq T - \beta\sqrt{n}} \binom{n}{r} \delta^r (1-\delta)^{n-r} S_n^r + \sum_{T-\beta\sqrt{n} < r \leq T} \binom{n}{r} \delta^r (1-\delta)^{n-r} S_n^r \\ &\leq S_n^{T-\beta\sqrt{n}} \sum_{r \leq T-\beta\sqrt{n}} \binom{n}{r} \delta^r (1-\delta)^{n-r} + S_n^T \sum_{T-\beta\sqrt{n} < r \leq T} \binom{n}{r} \delta^r (1-\delta)^{n-r} \\ &\overset{(35)}{\leq} S_n^{T-\beta\sqrt{n}} + S_n^T \beta\theta. \end{aligned} \tag{36}$$

In order to find an upper bound for $S$, it then suffices to find an upper bound for the right hand side in (36). For $1 \leq i \leq n$, let $Y_i \sim$ Bernoulli$(\frac{1}{2})$ and $X_i = \frac{1}{2} - Y_i$. Note that $\mathbb{E}[X_i] = 0$, $\mathbb{V}ar[X_i] = \frac{1}{4}$ and $|X_i| \leq \frac{1}{2}$. Then Theorem 12 implies

$$S_n^r = \Pr\left\{ \sum_{i=1}^n Y_i \leq r \right\} = \Pr\left\{ \sum_{i=1}^n X_i \geq \frac{n}{2} - r \right\} \leq \left( \frac{1}{2\sqrt{2\pi}} \cdot \frac{\sqrt{n}}{\frac{n}{2} - r} + \frac{\gamma}{\sqrt{n}} \right) e^{-nH(\frac{1}{4}, \frac{1}{2}, \frac{1}{2} - \frac{r}{n})}.$$

A simple calculation shows that $H(\frac{1}{4}, \frac{1}{2}, \frac{1}{2} - \frac{r}{n}) = \left(1 - h_2\left(\frac{r}{n}\right)\right) \ln 2$. Thus,

$$S_n^r \leq \left( \frac{1}{2\sqrt{2\pi}} \cdot \frac{1}{\frac{\sqrt{n}}{2} - \frac{r}{\sqrt{n}}} + \frac{\gamma}{\sqrt{n}} \right) 2^{-n\left(1 - h_2\left(\frac{r}{n}\right)\right)}. \tag{37}$$

Due to Lemma 1, part (i) and Lemma 2, part (ii), the first term in the right hand side of (37), when $r = T$, can be written as

$$\begin{aligned} \left( \frac{\sqrt{n}}{2} - \frac{T}{\sqrt{n}} \right)^{-1} &= \left( \sqrt{n} \left( \frac{1}{2} - \delta \right) - \sqrt{\delta(1-\delta)} Q^{-1} \left( p_e - \frac{p_e}{\sqrt{\kappa}} - \frac{c_1}{\sqrt{n}} \right) \right)^{-1} \\ &= \left( \sqrt{\frac{\ln 2}{2} \kappa} + \mathcal{O}\left( \frac{\kappa\sqrt{\kappa}}{n} \right) - \sqrt{\delta(1-\delta)} Q^{-1} \left( p_e - \frac{p_e}{\sqrt{\kappa}} - \frac{c_1}{\sqrt{n}} \right) \right)^{-1} \\ &= \left( \sqrt{\frac{\ln 2}{2} \kappa} \right)^{-1} + \mathcal{O}\left( \frac{\sqrt{\kappa}}{n} \right) + \mathcal{O}\left( \frac{\sqrt{-\log p_e}}{\kappa} \right) \\ &= \left( \sqrt{\frac{\ln 2}{2} \kappa} \right)^{-1} + \mathcal{O}\left( \frac{\sqrt{-\log p_e}}{\kappa} \right). \end{aligned} \tag{38}$$

Similarly when $r = T - \beta\sqrt{n}$, we have

$$\left(\frac{\sqrt{n}}{2} - \frac{T-\beta\sqrt{n}}{\sqrt{n}}\right)^{-1} = \left(\frac{\sqrt{n}}{2} - \frac{T}{\sqrt{n}} + \beta\right)^{-1} = \left(\sqrt{\frac{\ln 2}{2}\,\kappa}\right)^{-1} + \mathcal{O}\left(\frac{\sqrt{-\log p_e}}{\kappa}\right). \tag{39}$$

Now, the goal is to estimate the term $n\left(1 - h_2(\frac{r}{n})\right)$ in the right hand side of (37) for $r = T$ and $r = T - \beta\sqrt{n}$. Using the third order estimation of $h_2(.)$ gives

$$h_2\left(\frac{T}{n}\right) = h_2(\delta) + \left(\log_2 \frac{1-\delta}{\delta}\right)\sqrt{\frac{\delta(1-\delta)}{n}}Q^{-1}\left(p_e - \frac{p_e}{\sqrt{\kappa}} - \frac{c_1}{\sqrt{n}}\right) - \frac{1}{n\ln 2}Q^{-1}(p_e)^2 + \mathcal{O}\left(\frac{1}{n\sqrt{\kappa}}\right).$$

Therefore, by the definition of $\kappa$ and Lemma 1, part (ii), we obtain

$$n\left(1 - h_2\left(\frac{T}{n}\right)\right) = \kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}}\cdot\sqrt{\kappa}\,Q^{-1}\left(p_e - \frac{p_e}{\sqrt{\kappa}} - \frac{c_1}{\sqrt{n}}\right) + \frac{1}{\ln 2}Q^{-1}(p_e)^2 + \mathcal{O}\left(\frac{1}{\sqrt{\kappa}}\right).$$

Note that

$$\frac{d}{dx}Q^{-1}(x) = -\sqrt{2\pi}\,e^{\frac{Q^{-1}(x)^2}{2}}.$$

Thus, by applying Taylor expansion of $Q^{-1}(.)$, and using Lemma 2 parts (i) and (ii), we then conclude that

$$n\left(1 - h_2\left(\frac{T}{n}\right)\right) = \kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}}\cdot\sqrt{\kappa}\,Q^{-1}(p_e) + \mathcal{E}_1, \tag{40}$$

where

$$\mathcal{E}_1 = \frac{1}{\ln 2}Q^{-1}(p_e)^2 + \mathcal{O}\left(\frac{1}{\sqrt{-\log p_e}}\right) + \mathcal{O}\left(\frac{1}{\sqrt{\kappa}}\right) = \frac{1}{\ln 2}Q^{-1}(p_e)^2 + \mathcal{O}\left(\frac{1}{\sqrt{-\log p_e}}\right). \tag{41}$$

Exploiting the same analogy leads to the following result for $r = T - \beta\sqrt{n}$ :

$$\begin{aligned} n\left(1 - h_2\left(\frac{T-\beta\sqrt{n}}{n}\right)\right) &= \kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}}\cdot\sqrt{\kappa}\,Q^{-1}(p_e) + 2\sqrt{\frac{2}{\ln 2}}\cdot\beta\sqrt{\kappa} + \mathcal{E}_1 \\ &= \kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}}\cdot\sqrt{\kappa}\,Q^{-1}(p_e) + \frac{1}{2}\log_2\kappa + \mathcal{E}_1. \end{aligned} \tag{42}$$

Now, (37), (38), and (40) together imply

$$S_n^T \le \frac{1}{2\sqrt{\pi\kappa\ln 2}}2^{-\left(\kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}}\cdot\sqrt{\kappa}\,Q^{-1}(p_e) + \mathcal{E}_1\right)} + \mathcal{O}\left(\frac{\sqrt{-\log p_e}}{\kappa\,2^\kappa}\right). \tag{43}$$

Similarly, (37)), (39), and (42) together imply

$$S_n^{T-\beta\sqrt{n}} \le \frac{1}{2\sqrt{\pi\kappa\ln 2}}2^{-\left(\kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}}\cdot\sqrt{\kappa}\,Q^{-1}(p_e) + \frac{1}{2}\log_2\kappa + \mathcal{E}_1\right)} + \mathcal{O}\left(\frac{\sqrt{-\log p_e}}{\kappa\,2^\kappa}\right). \tag{44}$$

As a result, from (43), (44), and (36), we have

$$\begin{aligned} S &\le S_n^{T-\beta\sqrt{n}} + S_n^T\beta\theta \\ &\le \frac{1}{2\sqrt{\pi\ln 2}}\cdot\frac{1}{\kappa}\,(1 + \theta\log_2\kappa)\,2^{-\left(\kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}}\cdot\sqrt{\kappa}\,Q^{-1}(p_e) + \mathcal{E}_1\right)} + \mathcal{O}\left(\frac{\sqrt{-\log p_e}}{\kappa\,2^\kappa}\right). \end{aligned}$$

Now, taking the logarithm of both sides gives

$$-\log_2 S \geq \kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}} \cdot \sqrt{\kappa}\, Q^{-1}\left(p_e\right) + \mathcal{E}_1 + \log_2 \kappa - \log_2 \frac{1+\theta \log_2 \kappa}{2\sqrt{\pi \ln 2}} + \mathcal{O}\left(\frac{\sqrt{-\log p_e}}{\log \kappa}\right).$$

Therefore, by replacing $\mathcal{E}_1$ from (41), we find

$$-\log_2 S \geq \kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}} \cdot \sqrt{\kappa}\, Q^{-1}\left(p_e\right) + \log_2 \kappa + \frac{1}{\ln 2} Q^{-1}\left(p_e\right)^2 + \mathcal{O}\left(\log\log \kappa\right) + \mathcal{O}\left(\frac{1}{\sqrt{-\log p_e}}\right) + \mathcal{O}\left(\frac{\sqrt{-\log p_e}}{\log \kappa}\right).$$

Comparing the orders, then results in

$$-\log_2 S \geq \kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}} \cdot \sqrt{\kappa}\, Q^{-1}\left(p_e\right) + \log_2 \kappa + \frac{1}{\ln 2} Q^{-1}\left(p_e\right)^2 + \mathcal{O}\left(\log\log \kappa\right). \tag{45}$$

Finally, from (34) and (45), we conclude that

$$\begin{aligned}\log_2 M^*(n, p_e) &\geq \log_2 p_e - \frac{1}{2}\log_2 \kappa - \log_2 S \\ &\geq \kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}} \cdot \sqrt{\kappa}\, Q^{-1}\left(p_e\right) + \frac{1}{2}\log_2 \kappa + \log_2 p_e + \frac{1}{\ln 2} Q^{-1}\left(p_e\right)^2 + \mathcal{O}\left(\log\log \kappa\right).\end{aligned}$$

Note that by Lemma 2, part (ii), we have

$$\frac{1}{\ln 2} Q^{-1}\left(p_e\right)^2 = -2\log_2 p_e + \mathcal{O}(1).$$

Hence,

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

□

**Proof of Theorem 2. Converse Bound.** Let $X \sim$ Bernoulli($\frac{1}{2}$) and $Y$ be the input and output of the BSC($\delta$), respectively. Also suppose $P_X$, $P_Y$ and $P_{XY}$ are distributions of $X$, $Y$, and the joint distribution of $(X, Y)$ respectively. Define $P = P_{XY}$ and $Q = P_X P_Y$, and then define $P^n$ and $Q^n$ in terms of $P$ and $Q$ as they are in Lemma 4. Also consider $\beta_{1-p_e}(P^n, Q^n)$ as in (53). Under these choices of $P$ and $Q$, it can be verified that $\beta^n_{1-p_e}$ defined in (49), is a piecewise linear approximation of $\beta_{1-p_e}(P^n, Q^n)$ based on discrete values of error probabilities. Therefore, from (49), we can write

$$\log_2 M^*(n, p_e) \leq -\log_2 \beta_{1-p_e}(P^n, Q^n).$$

Now, using Lemma 4 for any $\gamma > 0$, we have

$$\log_2 M^*(n, p_e) \leq nD - \sqrt{nV} Q^{-1}\left(p_e + \gamma + \frac{B}{\sqrt{n}}\right) - \log_2 \gamma. \tag{46}$$

Under the specific values of $P$ and $Q$ given above, the quantities $D$,$V$,$T$, and $B$ in Lemma 4 can be computed as follows:

$$\begin{aligned} D &= 1 - h_2(\delta), \\ V &= \delta(1-\delta)\left(\log_2 \frac{1-\delta}{\delta}\right)^2, \\ T &= \delta(1-\delta)\left(\log_2 \frac{1-\delta}{\delta}\right)^3 \left(2\delta^2 - 2\delta + 1\right), \\ B &= \frac{2\delta^2 - 2\delta + 1}{\sqrt{\delta(1-\delta)}}. \end{aligned}$$

Replacing these quantities and using Lemma 1, we can rewrite (46) as the following:

$$\log_2 M^*(n, p_e) \leq \kappa - 2\sqrt{\frac{2\delta(1-\delta)}{\ln 2}} \cdot \sqrt{\kappa}\, Q^{-1}\left(p_e + \gamma + \frac{B}{\sqrt{n}}\right) - \log_2 \gamma + \mathcal{O}\left(\frac{\kappa\sqrt{-\kappa \log_2 p_e}}{n}\right).$$

Note that

$$\frac{d}{dx} Q^{-1}(x) = -\sqrt{2\pi}\, e^{\frac{Q^{-1}(x)^2}{2}}.$$

Therefore, by choosing $\gamma = \frac{p_e}{\sqrt{\kappa}}$ and applying the Taylor expansion of $Q^{-1}(.)$, and using Lemma 2, parts (i) and (ii), we can conclude that

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

Hence, by comparing the orders, we have

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

□

**Proof of Corollary 1.** In order to obtaining the optimal blocklength $n^*$ for transmission of $k$ information bits over a low-capacity BSC($\delta$), it suffices to replcae $M^*(n^*, p_e) = k$. Then $n^*$ can be computed by solving (5). Replace $M^*(n^*, p_e) = k$ and $\kappa = n^*C$ in (5), where $C = 1 - h_2(\delta)$, to obtain

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

Define $x = \sqrt{n^*C}$, $a = \sqrt{\frac{2\delta(1-\delta)}{\ln 2}} Q^{-1}(p_e)$ and $b = k + \log_2 p_e + \mathcal{O}(\log \kappa)$. Thus, we have

$$x^2 - 2ax - b = 0.$$

Note that the answer will be $x = a + \sqrt{a^2 + b}$. Therefore,

$$\sqrt{n^*C} = a + \sqrt{a^2 + b}.$$

More simplifications are as follows:

$$
\begin{aligned}
n^* &= \frac{1}{C}\left(a+\sqrt{a^2+b}\right)^2 \\
&= \frac{1}{C}\left(2a^2+b+2a\left(\sqrt{k}+\mathcal{O}\left(\frac{-\log p_e}{k}\right)+\mathcal{O}\left(\frac{\log\kappa}{k}\right)\right)\right) \\
&= \frac{1}{C}\left(k+2a\sqrt{k}+2a^2+\log_2 p_e+\mathcal{O}(\log\kappa)\right). \qquad (47)
\end{aligned}
$$

Note that in low-capacity regime, under optimal blocklength, $\log\kappa \approx \log k$. Therefore, by substituting the values of $C$ and $a$ in (47), we obtain

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

□

**Theorem 11** (Berry-Esseen)**.** *For $1 \le i \le n$, let $X_i$ be i.i.d random variables with $\mu = \mathbb{E}[X_i]$, $\sigma^2 = \mathbb{V}ar[X_i]$ and $\rho = \mathbb{E}\left[|X_i-\mu|^3\right] < \infty$. Then,*

$$
\left|\Pr\left\{\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma\sqrt{n}} > x\right\} - Q(x)\right| \le \frac{3\rho}{\sigma^3\sqrt{n}}.
$$

*Proof.* See [31, Theorem 3.4.9]. □

**Theorem 12** (A Sharp Tail Inequality)**.** *For $1 \le i \le n$, let $X_i$ be independent centered random variables such that $|X_i| \le b$. Define $\sigma^2 = \mathbb{V}ar\left[\sum_{i=1}^{n} X_i\right]$. Then, for $0 \le t \le \frac{\sigma}{\gamma b}$, we have*

$$
\Pr\left\{\sum_{i=1}^{n} X_i \ge t\right\} \le \left(\frac{1}{\sqrt{2\pi}}\frac{\sigma}{t}+\gamma\frac{b}{\sigma}\right)e^{-nH(\frac{\sigma^2}{n},b,\frac{t}{n})},
$$

*where*

$$
H(\nu,b,x) = \left(1+\frac{bx}{\nu}\right)\frac{\nu}{b^2+\nu}\ln\left(1+\frac{bx}{\nu^2}\right)+\left(1-\frac{x}{b}\right)\frac{b^2}{b^2+\nu}\ln\left(1-\frac{x}{b}\right).
$$

*Proof.* See [32, Theorem 1.1]. □

**Theorem 13** (RCU Achievability Bound for BSC)**.** *There exists an $(M, p_e)$-code over BSC$^n(\delta)$ such that*

$$
p_e \le \sum_{r=0}^{n}\binom{n}{r}\delta^r(1-\delta)^{n-r}\min\left\{1,(M-1)S_n^r\right\}, \qquad (48)
$$

*where*

$$
S_n^r = \sum_{s=0}^{r}\binom{n}{s}2^{-n}.
$$

*Proof.* See [12, Corollary 39]. □

**Theorem 14** (Converse Bound for BSC)**.** *For any $(M, p_e)$-code over the BEC$^n(\delta)$, we have*

$$M \leq \frac{1}{\beta^n_{1-p_e}}. \tag{49}$$

*where $\beta^n_\alpha$ for a real $\alpha \in [0, 1]$ is defined below based on values of $\beta_\ell$ where $\ell$ is an integer:*

$$\beta^n_\alpha = (1-\lambda)\beta_L + \lambda\beta_{L+1},$$
$$\beta_\ell = \sum_{r=0}^{\ell} \binom{n}{r} 2^{-n},$$

*such that $\lambda \in [0, 1)$ and nteger $L$ satisfy the following:*

$$\alpha = (1-\lambda)\alpha_L + \lambda\alpha_{L+1},$$
$$\alpha_\ell = \sum_{r=0}^{\ell-1} \binom{n}{r} \delta^r (1-\delta)^{n-r}.$$

*Proof.* See [12, Theorem 40]. □

**Lemma 1.** *Consider transmission over BSC($\delta$) in low-capacity regime and let $\kappa = n(1-h_2(\delta))$. Then the following hold:*

*i)* $\sqrt{n}\left(\frac{1}{2}-\delta\right) = \sqrt{\frac{\ln 2}{2}\,\kappa} + \mathcal{O}\left(\frac{\kappa\sqrt{\kappa}}{n}\right)$.

*ii)* $\sqrt{n}\,\log_2 \frac{1-\delta}{\delta} = 2\sqrt{\frac{2}{\ln 2}\,\kappa} + \mathcal{O}\left(\frac{\kappa\sqrt{\kappa}}{n}\right)$.

*Proof.* i) Taylor expansion of $h_2(\delta)$ around $\frac{1}{2}$ is given by

$$h_2(\delta) = 1 - \frac{1}{2\ln 2}\sum_{n=1}^{\infty} \frac{(1-2\delta)^{2n}}{n(2n-1)}.$$

Thus, the estimation of $h_2(\delta)$ up to the third order will be the following:

$$h_2(\delta) = 1 - \frac{1}{2\ln 2}(1-2\delta)^2 - \frac{1}{2\ln 2}\cdot\frac{1}{6}(1-2\delta)^4 + \mathcal{O}\left((1-2\delta)^6\right).$$

Therefore,

$$C = 1 - h_2(\delta) = \frac{1}{2\ln 2}(1-2\delta)^2 + \frac{1}{2\ln 2}\cdot\frac{1}{6}(1-2\delta)^4 + \mathcal{O}\left((1-2\delta)^6\right).$$

Now assuming $x = (1-2\delta)^2$ leads to the following equation:

$$\frac{x^2}{6} + x - 2C\ln 2 = 0.$$

Solving this equation gives

$$x = 3\left(-1 + \sqrt{1 + \frac{4}{3}C\ln 2}\right) = 2C\ln 2 + \mathcal{O}\left(C^2\right).$$

Therefore,

$$\frac{1}{2}-\delta=\frac{\sqrt{x}}{2}=\sqrt{\frac{\ln 2}{2}C}+\mathcal{O}\left(C^{\frac{3}{2}}\right).$$

Finally,

$$\sqrt{n}\left(\frac{1}{2}-\delta\right)=\sqrt{\frac{\ln 2}{2}nC}+\mathcal{O}\left(nC^{\frac{3}{2}}\right).$$

Note that $\kappa=nC$. As a result,

$$\sqrt{n}\left(\frac{1}{2}-\delta\right)=\sqrt{\frac{\ln 2}{2}\kappa}+\mathcal{O}\left(\frac{\kappa\sqrt{\kappa}}{n}\right). \tag{50}$$

ii) The first order estimation of the function $\log_2\frac{1-\delta}{\delta}$ around $\frac{1}{2}$ is

$$\sqrt{n}\,\log_2\frac{1-\delta}{\delta}=\frac{4}{\ln 2}\left(\frac{1}{2}-\delta\right)+\mathcal{O}\left(\left(\frac{1}{2}-\delta\right)^2\right).$$

Hence, using (50), we arrive at

$$\sqrt{n}\,\log_2\frac{1-\delta}{\delta}=2\sqrt{\frac{2}{\ln 2}\kappa}+\mathcal{O}\left(\frac{\kappa\sqrt{\kappa}}{n}\right).$$

□

**Lemma 2.** *Suppose* $0<p<Q(1)\approx 0.16$.[5] *Then the following hold:*

*(i)*

$$\sqrt{2\pi}\,p\,Q^{-1}(p)<e^{-\frac{Q^{-1}(p)^2}{2}}<\sqrt{2\pi}\,p\,\left(Q^{-1}(p)+1\right).$$

*(ii)*

$$\sqrt{8\pi+2-2\ln p}-2\sqrt{2\pi}<Q^{-1}(p)<\sqrt{-\ln 2\pi-2\ln p}.$$

*Proof.* i) For $x>0$, it is well known that

$$\frac{\phi(x)}{x+\frac{1}{x}}=\left(\frac{x}{1+x^2}\right)\phi(x)<Q(x)<\frac{\phi(x)}{x}. \tag{51}$$

Note that for $x>1$, (51) becomes

$$\frac{\phi(x)}{x+1}<Q(x)<\frac{\phi(x)}{x}. \tag{52}$$

Now, define $p=Q(x)$. Thus, $x=Q^{-1}(p)>1$. Therefore,

$$\frac{e^{-\frac{Q^{-1}(p)^2}{2}}}{\sqrt{2\pi}\,\left(Q^{-1}(p)+1\right)}<p<\frac{e^{-\frac{Q^{-1}(p)^2}{2}}}{\sqrt{2\pi}\,Q^{-1}(p)}.$$

Simplify to get

$$\sqrt{2\pi}\,p\,Q^{-1}(p)<e^{-\frac{Q^{-1}(p)^2}{2}}<\sqrt{2\pi}\,p\,\left(Q^{-1}(p)+1\right).$$

[5] Note that similar results can be obtained when $0<p<1$ but taking the assumption $0<p<Q(1)\approx 0.16$ into account leads to a very simple format for bounds and will not cause any conflict because we are interested in small bit and block error probabilities. More specifically, typical values of $p$ would be around $10^{-3}-10^{-6}$.

ii) From (52), for $x > 1$ we have

$$Q(x) < \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}.$$

Therefore,

$$x < \sqrt{-2\ln\left(\sqrt{2\pi}Q(x)\right)}.$$

Put $x = Q^{-1}(p)$ to get

$$Q^{-1}(p) < \sqrt{2\ln\frac{1}{\sqrt{2\pi}\,p}} = \sqrt{-\ln 2\pi - 2\ln p}.$$

For the other side, note that from (52), for $x > 1$, we also have

$$Q(x) > \frac{\phi(x)}{2x} = \frac{e^{-\frac{x^2}{2}}}{2x\sqrt{2\pi}}.$$

Assume $p = Q(x)$. As a result,

$$\frac{x^2}{2} + 2\sqrt{2\pi}\,x + \ln p - 1 > 0.$$

Hence,

$$Q^{-1}(p) = x > -2\sqrt{2\pi} + \sqrt{8\pi + 2 - 2\ln p}.$$

□

**Lemma 3.** *Suppose* $0 < \delta < 1$ *such that* $n\delta$ *is an integer. Then, for all* $0 \le r \le n$*, we have*

$$\binom{n}{r}\delta^r(1-\delta)^{n-r} \le \frac{\theta}{\sqrt{n}}.$$

*where*

$$\theta = \frac{e}{2\pi\sqrt{\delta(1-\delta)}}.$$

*Proof.* For $0 \le r \le n$ define

$$A(r) = \binom{n}{r}\delta^r(1-\delta)^{n-r},$$

and

$$r^* = \arg\max_{0\le r\le n} A(r).$$

First of all, note that from the Mode of Binomial distribution, we know that $r^* = \lfloor (n+1)\delta \rfloor = n\delta$. Also from Stirling formula, for any integer $n$, we have

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \le n! \le e\sqrt{n}\left(\frac{n}{e}\right)^n.$$

Therefore,

$$\begin{aligned} A(r^*) &= \binom{n}{n\delta}\delta^{n\delta}(1-\delta)^{n-n\delta} = \frac{n!\,\delta^{n\delta}(1-\delta)^{n-n\delta}}{n\delta!\,(n-n\delta)!} \\ &\le \frac{e\sqrt{n}\left(\frac{n}{e}\right)^n\delta^{n\delta}(1-\delta)^{n-n\delta}}{\sqrt{2\pi n\delta}\left(\frac{n\delta}{e}\right)^{n\delta}\sqrt{2\pi n(1-\delta)}\left(\frac{n(1-\delta)}{e}\right)^{n(1-\delta)}} \\ &= \frac{e}{2\pi\sqrt{\delta(1-\delta)}}\cdot\frac{1}{\sqrt{n}}. \end{aligned}$$

□

**Definition 1.** *Consider the following binary hypothesis test:*

$$H_0 : X \sim P,$$
$$H_1 : X \sim Q,$$

*where $P$ and $Q$ are two probability distributions on the same space $\mathcal{X}$. Suppose a continuous decision rule $\zeta$ as a mapping from observation space $\mathcal{X}$ to $[0,1]$ with $\zeta(x) \approx 0$ correspoding to reject $H_1$ and $\zeta(x) \approx 1$ correspoding to reject $H_0$. Now, define the smallest type-II error in this binary hypothesis test given that type-I error is not greater than $p_e$ as the following:*

$$\beta_{1-p_e}(P,Q) = \inf_{\zeta:\mathcal{X}\to[0,1]} \Big\{ \mathbf{E}_Q[1-\zeta(X)] : \mathbf{E}_P[\zeta(X)] \le p_e \Big\}. \tag{53}$$

**Lemma 4.** *Consider two discrete probability distributions $P$ and $Q$ on $\mathcal{X}$. Define the product distributions $P^n$ and $Q^n$ as*

$$P^n(\mathbf{x}) = \prod_{i=1}^{n} P(x_i), \qquad Q^n(\mathbf{x}) = \prod_{i=1}^{n} Q(x_i),$$

*where $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$. Then, for $p_e \in (0,1)$ and any $\gamma$, we have*

$$\log_2 \beta_{1-p_e}(P^n, Q^n) \ge -nD + \sqrt{nV} Q^{-1}\left(p_e + \gamma + \frac{B}{\sqrt{n}}\right) + \log_2 \gamma,$$

*where*

$$D = D(P\|Q),$$
$$V = \int_{\mathcal{X}} \left(\log_2 \frac{dP(x)}{dQ(x)} - D\right)^2 dP(x),$$
$$T = \int_{\mathcal{X}} \left|\log_2 \frac{dP(x)}{dQ(x)} - D\right|^3 dP(x),$$
$$B = \frac{T}{V^{\frac{3}{2}}}.$$

*Proof.* In the proof of [12, Lemma 14], put $\alpha = 1 - p_e, \Delta = \gamma\sqrt{n}, P_i = P, Q_i = Q$ and consider the logarithm in base 2. □

### 5.4 Proofs for AWGN Channel

In this section we will prove the converse and achievability bounds of Theorem 3. In the proofs we will be using Theorems 15–16 as well as Lemma 5 which are stated at the end of this section. For results in coding theory, we generally refer to [12] as it has well collected and presented the corresponding proofs. See also [8], [24], [25], [26], and [27].

**Proof of Theorem 3. Converse Bound.** Let $X$ and $Y$ be a uniform input and the corresponding output of an AWGN($\eta$) channel. Under the notation of Theorem 15, define $P = P_{XY}$ and $Q = P_X P_Y$. Therefore, we have

$$M^*(n, p_e) \le \frac{1}{\beta_{1-p_e}(P^n, Q^n)}. \tag{54}$$

Also from Lemma 4, we obtain

$$\log_2 \beta_{1-p_e}(P^n, Q^n) \geq -nD + \sqrt{nV} Q^{-1}\left(p_e + \gamma + \frac{B}{\sqrt{n}}\right) + \log_2 \gamma, \tag{55}$$

where $\gamma > 0$ and in order to compute the quantities $D$, $V$ and $B$, consider the random variable $H_n = \log_2 \frac{dP^n(z)}{dQ^n(z)}$. It can be verified that $H_n = \sum_{i=1}^{n} h_i$, where

$$h_i = \frac{1}{2}\log_2(1+\eta) + \frac{\eta}{2(1+\eta)\ln 2} - \frac{1}{2(1+\eta)\ln 2}\left(\eta Z_i^2 - 2\sqrt{\eta}\, Z_i\right),$$

assuming that $Z_i$'s are independent standard normal random variables. Thus, a simple calculation shows that for all $i \in \{1, \ldots, n\}$, we have

$$\begin{aligned}
D &= \mathbb{E}[h_i] = \frac{1}{2}\log_2(1+\eta), \\
V &= \mathbb{V}ar[h_i] = \frac{\eta(\eta+2)}{2(\eta+1)^2 \ln^2 2}, \\
T &= \mathbb{E}\left[|h_i - D|^3\right], \\
B &= \frac{T}{V^{\frac{3}{2}}}.
\end{aligned}$$

As we will see in the rest of the proof, computing $T$ and $B$ are not necessary as they will appear in terms which vanish compared to other terms. Now, by replacing the values computed above, (54) together with (55) yield

$$\log_2 M^*(n, p_e) \leq \frac{n}{2}\log_2(1+\eta) - \frac{\sqrt{n\eta(\eta+2)}}{\sqrt{2}(\eta+1)\ln 2} Q^{-1}\left(p_e + \gamma + \frac{B}{\sqrt{n}}\right) - \log_2 \gamma.$$

Now, by setting $\gamma = \frac{p_e}{\sqrt{\kappa}}$ and plugging in Lemma 5, and then using Lemma 2, part (ii) together with the definition of $\kappa$, we conclude that

$$\log_2 M^*(n, p_e) \leq \kappa - \frac{\sqrt{\kappa(\eta+2)}}{(\eta+1)\sqrt{\ln 2}} Q^{-1}\left(p_e + \frac{p_e}{\sqrt{\kappa}} + \frac{B}{\sqrt{n}}\right) - \log_2 \frac{p_e}{\sqrt{\kappa}} + \mathcal{O}\left(\frac{\kappa^2\sqrt{-\log p_e}}{n}\right).$$

Note that

$$\frac{d}{dx} Q^{-1}(x) = -\sqrt{2\pi}\, e^{\frac{Q^{-1}(x)^2}{2}}.$$

Therefore, by applying the Taylor expansion of $Q^{-1}(.)$, and using Lemma 2 parts (i) and (ii), we arrive at

$$\log_2 M^*(n, p_e) \leq \kappa - \frac{\sqrt{\eta+2}}{(\eta+1)\sqrt{\ln 2}} \cdot \sqrt{\kappa}\, Q^{-1}(p_e) - \log_2 \frac{p_e}{\sqrt{\kappa}} + \mathcal{O}\left(\frac{\kappa^2\sqrt{-\log_2 p_e}}{n}\right) + \mathcal{O}\left(\frac{1}{\sqrt{-\log p_e}}\right).$$

Hence, comparing the orders, results in

$$\log_2 M^*(n, p_e) \leq \kappa - \frac{\sqrt{\eta+2}}{(\eta+1)\sqrt{\ln 2}} \cdot \sqrt{\kappa}\, Q^{-1}(p_e) + \frac{1}{2}\log_2 \kappa - \log_2 p_e + \mathcal{O}\left(\frac{1}{\sqrt{-\log p_e}}\right).$$

□

**Proof of Theorem 3. Achievability Bound.** The proof of Theorem 16 is still valid in low-capacity regime. As a result, using Theorem 16, together with Lemma 5 and Lemma 2, part (ii), yield

$$\begin{aligned}\log_2 M^*(n, p_e, \eta) &\geq \kappa - \frac{\sqrt{\eta+2}}{(\eta+1)\sqrt{\ln 2}} \cdot \sqrt{\kappa}\, Q^{-1}(p_e) + \mathcal{O}(1) + \mathcal{O}\left(\frac{\kappa\sqrt{-\kappa \log p_e}}{n}\right) \\ &= \kappa - \frac{\sqrt{\eta+2}}{(\eta+1)\sqrt{\ln 2}} \cdot \sqrt{\kappa}\, Q^{-1}(p_e) + \mathcal{O}(1).\end{aligned}$$

□

**Proof of Corollary 2.** In order to(ii)ing the optimal blocklength $n^*$ for transmission of $k$ information bits over a low capacity AWGN($\eta$), it suffices to replcae $M^*(n^*, p_e, \eta) = k$. Then $n^*$ can be computed by solving (6). Substitute $M^*(n^*, p_e) = k$ and $\kappa = n^* C$ in (6), where $C = \frac{1}{2}\log_2(1+\eta)$, to obtain

$$k = n^* C - \frac{\sqrt{\eta+2}}{(\eta+1)\sqrt{\ln 2}} \cdot \sqrt{n^* C}\, Q^{-1}(p_e) + \mathcal{E}.$$

Define $x = \sqrt{n^* C}$, $a = \frac{\sqrt{\eta+2}}{2(\eta+1)\sqrt{\ln 2}} Q^{-1}(p_e)$ and $b = k - \mathcal{E}$. Thus, we have

$$x^2 - 2ax - b = 0.$$

Note that the answer will be $x = a + \sqrt{a^2+b}$. Therefore,

$$\sqrt{n^* C} = a + \sqrt{a^2+b}.$$

More simplifications are as follows:

$$\begin{aligned}n^* &= \frac{1}{C}\left(a + \sqrt{a^2+b}\right)^2 \\ &= \frac{1}{C}\left(2a^2 + b + 2a\left(\sqrt{k} + \mathcal{O}\left(\frac{-\log p_e}{k}\right) + \mathcal{O}\left(\frac{\log \kappa}{k}\right)\right)\right) \\ &= \frac{1}{C}\left(k + 2a\sqrt{k} + 2a^2 - \mathcal{E} + \mathcal{O}\left(\frac{(-\log p_e)^{3/2}}{k}\right)\right). \qquad (56)\end{aligned}$$

Note that in low-capacity regime, under optimal blocklength, $\log \kappa \approx \log k$. Therefore, by substituting the values of $C$ and $a$ in (56) and comparing the orders, we obtain

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

Now, applying Lemma 2, part (ii), results in

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

□

**Theorem 15** (General Converse Bound)**.** *Consider $n$ independent uses of a channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. Let $P_X$, $P_Y$, and $P_{XY}$ be distributions on $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{X} \times \mathcal{Y}$ respectively and suppose $P_X^n$, $P_Y^n$, and $P_{XY}^n$ are their product distribution over $n$ independent trials. Then the following holds:*

$$M^*(n, p_e) \leq \sup_{P_X} \inf_{P_Y} \frac{1}{\beta_{1-p_e}\left(P_{XY}^n, P_X^n P_Y^n\right)}.$$

*where $\beta_{1-p_e}$ is defined as* (53).

*Proof.* See [12, Theorem 29]. □

**Theorem 16** (Achievability Bound for AWGN)**.** *There exists an $(M, p_e, \eta)$-code over AWGN$^n(\eta)$ such that*

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

*where*

$$C = \frac{1}{2}\log_2\left(1+\eta\right),$$
$$V = \frac{\eta(\eta+2)}{2(\eta+1)^2 \ln^2 2}.$$

*Proof.* See the achievability bound in [12, Theorem 73]. □

**Lemma 5.** *Consider transmission over AWGN($\eta$) in low-capacity regime where by definition $\kappa = \frac{n}{2}\log_2(1+\eta)$. Then we have*

$$n\eta = 2\,\kappa \ln 2 + \mathcal{O}\left(\frac{\kappa^2}{n}\right).$$

*Proof.* Consider $C = \frac{1}{2}\log(1+\eta)$. Thus, by using Taylor expansion of $\log(1+x)$ up to the second order, we arrive at

$$C = \frac{1}{2\ln 2}\ln(1+\eta) = \frac{1}{2\ln 2}\left(\eta - \frac{\eta^2}{2}\right) + \mathcal{O}\left(\eta^3\right).$$

which leads to solving

$$\eta^2 - 2\eta + 4C\ln 2 = 0.$$

As a result,

$$\eta = 1 - \sqrt{1 - 4C\ln 2} = 2\,C\ln 2 + \mathcal{O}\left(C^2\right).$$

Now, considering $\kappa = nC$, results in

$$n\eta = 2\,nC\ln 2 + \mathcal{O}\left(nC^2\right) = 2\,\kappa\ln 2 + \mathcal{O}\left(\frac{\kappa^2}{n}\right).$$

□

### 5.5 Proofs For Repetition Sufficiency

**Proof of Theorem 4.** Let $m$ be the number of repeated blocks of size $r$, i.e., $m = \frac{n}{r}$ and consequently, $m_\beta = \frac{n}{r_\beta}$. Note that the maximum achievable rate in this setting is $\frac{1-\epsilon^r}{r} = \frac{m}{n}\left(1-\epsilon^{\frac{n}{m}}\right)$. Thus, $m_\beta$ is the solution of the following problem:

$$\begin{aligned} &\text{minimize} \quad m \\ &\text{subject to} \quad m\left(1-\epsilon^{\frac{n}{m}}\right) \geq \beta n(1-\epsilon). \end{aligned}$$

It can be easily verified that the answer of this problem $m_\beta$ indeed satisfies

$$m_\beta\left(1-\epsilon^{\frac{n}{m_\beta}}\right) = \beta n(1-\epsilon).$$

Now putting $\kappa = n(1-\epsilon)$ gives

$$m_\beta\left(1-\epsilon^{\frac{\kappa}{(1-\epsilon)m_\beta}}\right) = \beta\kappa.$$

Define $x = \frac{\kappa}{m_\beta}$ to get

$$1-\epsilon^{\frac{x}{1-\epsilon}} = \beta x.$$

Therefore,

$$1-\beta x = \epsilon^{\frac{x}{1-\epsilon}} = e^{\frac{\ln\epsilon}{1-\epsilon}x} = e^{-x\ell},$$

where $\ell = -\frac{\ln\epsilon}{1-\epsilon}$. Now, let $z = x\ell$. Thus,

$$1-\frac{\beta}{\ell}z = e^{-z}.$$

Finally, define $\gamma = \frac{\beta}{\ell}$ to get

$$1-\gamma z = e^{-z}. \tag{57}$$

Then, using Lemma 6, part (ii), we can write

$$\begin{aligned} 1-\gamma z = e^{-z} &\leq 1-z+\frac{z^2}{2}, \\ -\gamma &\leq -1+\frac{z}{2}, \\ 2(1-\gamma) &\leq z. \end{aligned} \tag{58}$$

Taking logarithm of both sides of (57), gives

$$-z = \ln(1-\gamma z).$$

Now, use Lemma 6, part (iii) to obtain

$$-z = \ln(1-\gamma z) \leq -\gamma z - \frac{(\gamma z)^2}{2}.$$

Therefore,

$$z \leq 2(1-\gamma)\frac{1}{\gamma^2}. \tag{59}$$

Note that (58) together with (59) yields

$$2(1-\gamma) \leq z \leq 2(1-\gamma)\frac{1}{\gamma^2}.$$

Remember $z = x\ell = \frac{\kappa\ell}{m_\beta}$ . Hence,

$$\frac{\kappa\ell}{2(1-\gamma)} \cdot \gamma^2 \leq m_\beta \leq \frac{\kappa\ell}{2(1-\gamma)}.$$

Now, replacing $\gamma = \frac{\beta}{\ell}$, $\kappa = n(1-\epsilon)$, and $m_\beta = \frac{n}{r_\beta}$ result in

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

□

**Proof of Theorem 5.** We use extremes of information combining. Consider two BMS channels $W_1, W_2$ with capacity $C_1, C_2$ respectively. Note that $\text{BEC}(1-C_1)$ has capacity $C_1$ and $\text{BEC}(1-C_2)$ has capacity $C_2$. Then we know from extremes of information combining [33, Chapter 4] that

$$C(W_1 \circledast W_2) \leq C\left(\text{BEC}(1-C_1) \circledast \text{BEC}(1-C_2)\right), \tag{60}$$

where $W_1 \circledast W_2$ is the BMS channel whose out put is formed as the union of the output of $W_1$ and $W_2$, i.e. we send the input bit once through $W_1$ and once through $W_2$ and the resulting outcomes together will be the outcome of $W_1 \circledast W_2$.

Now, it is clear that for any BMS channel $W$ we have

$$W^r = \underbrace{W \circledast \cdots \circledast W}_{r \text{ times}}.$$

Thus, assuming $C(W) = C$, then by using (60) $r$ times we obtain $C(W^r) \leq C(\text{BEC}(1-C)^r)$. □

**Lemma 6.** *Suppose $z \geq 0$. Then,*

*i)* $e^{-z} \geq 1 - z$.

*ii)* $e^{-z} \leq 1 - z + \frac{z^2}{2}$.

*iii) If* $0 < z < 1$, *then* $\ln(1-z) \leq -z - \frac{z^2}{2}$.

*Proof.* i) Define $f(z) = e^{-z} - 1 + z$. Note that $f'(z) = -e^{-z} + 1 \geq 0$ for $z \geq 0$. This means $f$ is increasing over $z \geq 0$. Thus, if $z \geq 0$, then $f(z) \geq f(0) = 0$.

ii) Define $g(z) = e^{-z} - 1 + z - \frac{z^2}{2}$. Note that $g'(z) = -e^{-z} + 1 - z \leq 0$ for $z \geq 0$ due to part (i). This means $g$ is decreasing over $z \geq 0$. Thus, if $z \geq 0$, then $g(z) \leq g(0) = 0$.

iii) $\ln(1-z) = -\sum_{i=1}^{\infty} \frac{z^i}{i} \leq -z - \frac{z^2}{2}$ for $0 < z < 1$.

□

### 5.6 Proof of Theorem 6

Let $i \in [n]$ be an arbitrary index. Recall that in order to construct the $i$-th sub-channel for a polar code of length $n = 2^m$ on channel $W$, which is denoted by $W_n^{(i)}$, we proceed as follows: (i) Consider the binary expansion $i = b_1 b_2 \cdots b_m$. (ii) Start with $W_0 = W$. (iii) For $j \in \{1, \ldots, m\}$, let $W_j = W_{j-1} \circledast W_{j-1}$ if $b_j = 1$, and otherwise, let $W_j = W_{j-1} \boxdot W_{j-1}$. (iv) The channel $W_m$ is the sub-channel corresponding to the $i$-th index. Also recall that for any BMS channel $W$, we have (see [34, Lemma 3.16], [35])

$$Z(W \circledast W) = Z(W)^2 \quad \text{and} \quad Z(W \boxdot W) \geq \sqrt{1 - (1 - (1 - Z(W))^2)^2},$$

which by simple manipulations will be simplified to

$$1 - Z(W \circledast W) \leq 2(1 - Z(W)) \quad \text{and} \quad 1 - Z(W \boxdot W) \leq 4 * (1 - Z(W))^2. \tag{61}$$

Now, for an integer $t \leq m$ let $i_t$ be such that in the binary expansion $i_t = b_1 b_2 \cdots b_m$, all the bits $b_j$ are equal to 1 except for $b_t$ which is 0 (i.e., $i_t = 2^m - 2^{m-t} - 1$). Using the bounds in (61), we can write

$$1 - Z(W_n^{(i_t)}) \leq 2^{m-t+2} \left(2^{t-1}(1 - Z(W))\right)^2.$$

Thus, if $i_t$ is a good sub-channel, then we must have $Z(W_m^{(i_t)}) \leq \frac{1}{2}$ which by using the above inequality gives us

$$2^{m-t+2} \left(2^{t-1}(1 - Z(W))\right)^2 \geq \frac{1}{2} \Longrightarrow 2^t \geq \frac{1}{2^{m+1}(1 - Z(W))^2}.$$

Note also that for any BMS channel, we have $C(W) \geq 1 - Z(W)$, and thus the above inequality implies

$$2^t \geq \frac{1}{2^{m+1} C(W)^2} \Longrightarrow 2^{m-t+1} \leq 4(2^m C(W))^2 = 4(n(c(W)))^2. \tag{62}$$

Now, recalling the fact that the binary expansion of index $i_t$ has only one position with zero value (the $t$-th position), we can conclude that for any other good index $i$, we must have the following: The first $t-1$ bits of the binary expansion of $i-1$ should be 1, and $t$ is lower bounded from (62). This means that the polar code corresponding to $W$ will have at least $2^{t-1}$ repetitions in the beginning.

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

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

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

[6] 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.

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

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

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

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

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

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

[13] S. H. Hassani, S. B. Korada, and R. Urbanke, "The compound capacity of polar codes," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*. IEEE, 2009, pp. 16–21.

[14] E. Sasoglu and L. Wang, "Universal polarization," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 2937–2946, 2016.

[15] S. H. Hassani and R. Urbanke, "Universal polar codes," in *2014 IEEE International Symposium on Information Theory*. Ieee, 2014, pp. 1451–1455.

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

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

[18] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2213–2226, 2015.

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

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

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

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

[23] H. Mahdavifar, “Fast polarization for non-stationary channels,” in *Information Theory (ISIT), 2017 IEEE International Symposium on*. IEEE, 2017, pp. 849–853.

[24] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. New York: Springer-Verlag, 1978.

[25] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[26] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.

[27] P. Elias, “Coding for two noisy channels,” in *Proc. 3d London Symp. Inf. Theory. Washington, DC*, Sep. 1955, pp. 61–76.

[28] V. Zacharovas and H.-K. Hwang, “A charlier–parseval approach to poisson approximation and its applications,” *Lithuanian Mathematical Journal*, no. 1, pp. 88–119, 2010.

[29] C. Stein, *Approximate Computation of Expectations*. Inst. of Mathematical Statistic, 1987.

[30] M. D. Hirschhorn and M. B. Villarino, “The missing factor in hoeffding’s inequalities,” *The Ramanujan Journal*, no. 34, pp. 73–81, 2014.

[31] R. Durrett, *Probability: Theory and Examples*. Cambridge University Press, 2010.

[32] M. Talagrand, “The missing factor in hoeffding’s inequalities,” *Annales de l’I.H.P. Probabilités et statistiques*, no. 4, pp. 689–702, 1995.

[33] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[34] S. B. Korada, “Polar codes for channel and source coding,” Ph.D. dissertation, Ecole Polytechnique Federale de Lausanne, 7 2009.

[35] S. H. Hassani, “Polarization and spatial coupling: Two techniques to boost performance,” Ph.D. dissertation, Ecole Polytechnique Federale de Lausanne, 9 2013.