Near-Optimal Finite-Length Scaling for Polar Codes over Large Alphabets

Henry D. Pfister and Rüdiger Urbanke

Abstract—For any prime power $q$, Mori and Tanaka introduced a family of $q$-ary polar codes based on $q$ by $q$ Reed-Solomon polarization kernels. For transmission over a $q$-ary erasure channel, they also derived a closed-form recursion for the erasure probability of each effective channel. In this paper, we use that expression to analyze the finite-length scaling of these codes on $q$-ary erasure channel with erasure probability $\epsilon \in (0, 1)$. Our primary result is that, for any rate $R$, the blocklength must tend to infinity. A more refined question is, “How fast can the gap to capacity decrease as a function of the blocklength?” A key result is that, for any rate-$R$ code achieving a block error rate of $\eta < 1$ on a non-trivial discrete memoryless channel with capacity $C$, the blocklength $N$ must satisfy $C - R \geq A / \sqrt{N}$ for some $A > 0$ that depends only on $\delta$ and the channel [11]. Thus, the gap to capacity cannot vanish faster than $O(N^{-1/2})$. Random codes are known to achieve this scaling.

Polar codes are the first codes, with low-complexity encoding and decoding algorithms, that were proven to achieve capacity on binary-input memoryless channels [5], [6]. Since then, the rate of polarization and the relationship between the blocklength and the error rate has received significant attention [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. This relationship is typically studied in two distinct regimes by asking two different questions. First, for a fixed rate $R < C$, how fast does the error rate decay with the blocklength? Second, for a fixed probability of decoding failure $\eta \in (0, 1)$, how fast can the rate approach the capacity?

The majority of prior work in this area focuses on binary polar codes with $2 \times 2$ kernels and, for these codes, the gap to capacity cannot decrease faster than $O(N^{-0.276})$ [10], [16]. For $2 \times 2$ kernels with larger alphabets, the analysis is much more difficult and the provable scaling rates are even smaller [12], [15]. Recently, $8 \times 8$ and $16 \times 16$ binary kernels have been constructed that achieve scaling rates of $O(N^{-0.279})$ and $O(N^{-0.298})$ [11]. Until now, no reported results provably established scaling rates faster than $O(N^{-0.30})$.

In this work, we consider the $q$-ary polar codes introduced by Mori and Tanaka based on $q \times q$ Reed-Solomon (RS) polarization kernels with elements from the Galois field $\mathbb{F}_q$ [17], [18]. Thus, in all statements, $q$ is implicitly assumed to be a prime power. These codes have length $N = q^n$, where $n$ is the number of steps in the polarization process. We consider transmission over the $q$-ary erasure channel (QEC) with erasure probability $\epsilon$. Mori and Tanaka have also shown that these polar codes achieve capacity on symmetric $q$-ary channels [18].

By analyzing the polarization process for the QEC, we show that, for any $\gamma > 0$ and $\delta > 0$, there is a $q_0$ such that, for all $q \geq q_0$, the fraction of effective channels with erasure rate at most $N^{-\gamma}$ is at least $1 \geq 1 - \epsilon \geq O(N^{-1/2+\delta})$. Thus, the gap to capacity scales at a nearly-optimal rate. While our proof relies on large alphabet QECs with large polarization kernels, we believe a similar result may also hold for small alphabets (e.g., binary) with large polarization kernels.

Like binary polar codes, the performance of $q$-ary polar codes can be analyzed by tracking the evolution of the effective channels through the polarization process [6]. At each step, a single effective channel with erasure rate $\alpha$ splits into $q$ new channels. For their codes, Mori and Tanaka showed that the $i$-th new
effective channel, for \( i \in \mathbb{Q} \triangleq \{0, 1, \ldots, q - 1\} \), is a \( q \)-ary erasure channel with erasure probability

\[
\psi_i(x) = \sum_{j=1}^{q} \binom{q}{j} x^j (1 - x)^{q-j},
\]

(1)

Applying this formula recursively, one can compute the erasure rates of the \( N = q^n \) effective channels after \( n \) steps. For a polar code with \( k \) information symbols, the next step in the design process consists of choosing the \( k \) effective channels with the smallest erasure rates. For \( q = 2 \), these steps are identical to the original polar code construction in [6] and the resulting codes are closely related to binary Reed-Muller codes. For larger \( q \), the resulting codes are closely related to \( q \)-ary Reed-Muller codes [19].

II. THE POLARIZATION PROCESS

Let the random variable \( X_n \) denote the channel erasure probability for a randomly chosen effective channel after \( n \) levels of polarization. The sequence \( X_n \), for \( n = 0, 1, \ldots \), is a homogeneous Markov chain on the compact state space \( \mathcal{X} = [0, 1] \) with transition probability

\[
P\left( X_n = x_n \mid X_0, \ldots, X_{n-1} = (x_0, \ldots, x_{n-1}) \right)
= P( X_n = x_n \mid X_{n-1} = x_{n-1} ),
= \begin{cases} \frac{1}{q} & \text{if } \exists i \in \mathbb{Q}, \ x_{n-1} = \psi_i(x) \\ 0 & \text{otherwise.} \end{cases}
\]

We note that 0 and 1 are both absorbing states of this Markov chain and we are interested in the convergence rate to these states [6].

Let \( C(\mathcal{X}) \) denote the set of bounded continuous functions mapping \( \mathcal{X} \) to \( \mathbb{R} \). One can analyze this Markov chain by focusing on the sequence of functions, \( g_n(x) \triangleq \mathbb{E}[g_0(X_n) \mid X_0 = x] \), generated by \( g_0 \in C(\mathcal{X}) \) [10], [16]. Since the Markov chain is homogeneous, this sequence satisfies the recursion

\[
g_n(x) \triangleq \mathbb{E}[g_0(X_n) \mid X_0 = x]
= \sum_{i=0}^{q-1} \mathbb{E}[g(X_n) \mid X_1 = \psi_i(x)] P(X_1 = \psi_i(x) \mid X_0 = x)
= \sum_{i=0}^{q-1} g_{n-1}(\psi_i(x)) \frac{1}{q}.
\]

The one-step update is given by the linear operator \( T_q : C(\mathcal{X}) \rightarrow C(\mathcal{X}) \), which is defined by

\[
(T_q g_{n-1})(x) \triangleq \frac{1}{q} \sum_{i=0}^{q-1} g_{n-1}(\psi_i(x)).
\]

Since the polarization process preserves the average mutual information, it also preserves average erasure rate. This implies that the function \( g_0(x) = x \) should be an eigenfunction of \( T_q \) (with eigenvalue 1) and, using (2), one can verify that it is. We note that this is a straightforward generalization of the approach used for binary polar codes [10], [16].

The rate of polarization is determined by the fraction of channels whose erasure rates are not extremal. The following lemma connects the fraction of non-extremal channels (as a function of \( n \)) with an easily computable constant associated with \( T_q \). This can be seen as a standard convergence analysis based on Lyapunov functions and it was first applied to polar codes in [10].

Lemma 1. Suppose there exists a non-negative continuous function \( V : \mathcal{X} \rightarrow [0, \infty) \) and a constant \( \lambda \in (0, 1) \) such that

\[
(T_q V)(x) \leq \lambda V(x)
\]

(3)

for all \( x \in \mathcal{X} \). Then, for \( S(\alpha) \triangleq \{ x \in \mathcal{X} \mid V(x) \geq \alpha \} \), it follows that

\[
P(X_n \in S(\alpha) \mid X_0 = x) \leq \frac{\lambda^n V(x)}{\alpha}.
\]

Further, if \( S(\alpha) \) is a closed interval, then

\[
P(X_n \geq \min S(\alpha) \mid X_0 = x) \leq \frac{\lambda^n V(x)}{\alpha} + \frac{x}{\max S(\alpha)}.
\]

Proof: To see this, we choose \( g_0(x) = V(x) \) and observe that (3) implies \( \mathbb{E}[V(X_n) \mid X_0 = x] = g_n(x) \leq \lambda^n V(x) \) for all \( x \in \mathcal{X} \). From this, we get

\[
P(X_n \in S(\alpha) \mid X_0 = x) = P(V(X_n) \geq \alpha \mid X_0 = x)
\leq \frac{\mathbb{E}[V(X_n) \mid X_0 = x]}{\alpha}
\leq \frac{\lambda^n V(x)}{\alpha}.
\]

Since the polarization process preserves the average mutual information, we have \( \mathbb{E}[X_n \mid X_0 = x] = x \). For the second part, we combine this with the Markov inequality to see that

\[
P(X_n > \max S(\alpha) \mid X_0 = x) \leq \frac{\mathbb{E}[X_n \mid X_0 = x]}{\max S(\alpha)} = \frac{x}{\max S(\alpha)}.
\]

Since \( S(\alpha) \) is a closed interval, it follows that

\[
P(X_n \geq \min S(\alpha) \mid X_0 = x) = P(X_n \in S(\alpha) \mid X_0 = x)
+ P(X_n > \max S(\alpha) \mid X_0 = x)
\leq \frac{\lambda^n V(x)}{\alpha} + \frac{x}{\max S(\alpha)}.
\]

This completes the proof.
Remark 2. For the considered problem, this lemma is a slight variation of what is used in [8], [16]. We use this form to show the close connection to Lyapunov functions. From that perspective, the function \( V(x) \) can be seen as a Lyapunov function showing convergence to stationary distributions supported on the set \( \{ x \in X \mid V(x) = 0 \} \).

Definition 3. Let \( V(x) = (x(1-x))^\beta \) for \( \beta > 0 \) and define
\[
\lambda_{q,\beta} \triangleq \sup_{x \in (0,1)} \frac{(T_q V)(x)}{V(x)}.
\]

Then, \( \lambda_{q,\beta} \) is the largest \( \lambda \in \mathbb{R} \) such that \( (T_q V)(x) \leq \lambda V(x) \) for all \( x \in (0,1) \). We also note that \( V(x) \leq V \left( \frac{1}{2} \right) = (\frac{1}{4})^\beta \) for \( x \in [0,1] \).

Lemma 4. The quantity \( \lambda_{q,\beta} \) for \( \beta \in (0,\frac{1}{2}] \) satisfies
\[
\lambda_{q,\beta} \leq \frac{6}{\sqrt{q^\beta}} \left( \frac{1}{4} \right)^{\frac{1}{2} - \beta}.
\]

Proof: See Section III-C.

Corollary 5. If the conditions of Lemma 2 hold for \( V(x) = (x(1-x))^\beta \) with \( \beta > 0 \), then
\[
P(X_n \in [\eta,1-\eta]|X_0 = x) \leq \frac{\lambda^n V(x)}{V(\eta)}
\]
for \( \eta \in (0,\frac{1}{2}] \). This also implies
\[
P(X_n \geq \eta|X_0 = x) \leq \frac{\lambda^n V(x)}{V(\eta)} + \frac{x}{1-\eta}.
\]

Proof: The first statement follows from applying Lemma 2 with \( \alpha = V(\eta) \). For the second statement, we observe that \( S(\alpha) \) is a closed interval because \( V(x) \) is a concave function. Also, \( \max S(\alpha) = 1-\eta \) because \( V(\eta) = V(1-\eta) \) implies that \( 1-\eta \in S(\alpha) \) and \( V(x) < V(\eta) \) for \( x > 1-\eta \). This completes the proof.

The primary purpose of this paper is the statement and proof of the following theorem.

Theorem 6. For the q-ary polar codes defined in [17], [18], let \( X_N \) be the erasure rate of a randomly chosen effective channel after \( n \) steps of polarization. For any \( \gamma > 0 \), \( \beta \in (0,\frac{1}{2}] \), and \( N^{-\gamma} \leq \frac{3}{4} \), one finds that
\[
P(X_n \in [N^{-\gamma},1-N^{-\gamma})|X_0 = x) \leq N^{\gamma \beta \frac{1}{2} \ln \frac{6}{q^\beta} (1 - \frac{4}{1 + \beta})} \ln 4
\]

and
\[
P(X_n \geq N^{-\gamma} | X_0 = x) \leq N^{\gamma \beta \ln q^\beta (1 - \frac{4}{1 + \beta}) \ln 4} + \frac{x}{1-N^{-\gamma}}.
\]

Proof: Combining Lemma 4 and Corollary 5 with \( \eta = N^{-\gamma} \), one gets the prediction
\[
P(X_n \in [N^{-\gamma},1-N^{-\gamma})|X_0 = x) \leq \frac{(x(1-x))^\beta}{(N^{-\gamma}(1-N^{-\gamma}))^\beta} \left( 6 \sqrt{q^\beta} \left( \frac{1}{4} \right)^{\frac{1}{2} - \beta} \right)^n
\]
\[
\leq \left( \frac{1}{1-N^{-\gamma}} \right) \gamma^\beta q^\beta \ln \left( \frac{6}{q^\beta (1 - \frac{4}{1 + \beta})} \ln 4 \right)
\]
\[
\leq N^{\gamma \beta \frac{1}{2} \ln \frac{6}{q^\beta} (1 - \frac{4}{1 + \beta}) \ln 4},
\]

for \( N^{-\gamma} \leq \frac{3}{4} \). The second statement follows directly from the second part of Corollary 5.

Corollary 7. Consider the q-ary polar codes defined in [17], [18] on a QEC with erasure probability \( \gamma \). For any \( \gamma > 0 \) and \( \delta > 0 \), there is a \( \beta \in (0,\frac{1}{2}] \) and a \( q_0 \) such that, for all \( q \geq q_0 \), the fraction of effective channels with erasure rate at most \( N^{-\gamma} \) is at least \( 1 - \delta - O(N^{-1/2+\epsilon}) \).

Proof: Since the stated condition becomes weaker as \( \gamma \) decreases and \( \delta \) increases, we assume without loss of generality that \( \gamma \leq \frac{1}{2} \) and \( \delta \leq \frac{1}{4} \). Using this, we choose \( \beta = \frac{\delta}{2\gamma} \) and observe that \( \beta \in (0,\frac{1}{2}] \). At errror rate \( N^{-\gamma} \), the gap to capacity is given by
\[
(1 - \delta) - \frac{N^{\frac{1}{2} - \beta \ln q^\beta (1 - \frac{4}{1 + \beta}) \ln 4 + \epsilon N^{-\gamma}}}{1-N^{-\gamma}}
\]
\[
\leq N^{\frac{1}{2} - \beta \ln q^\beta (1 - \frac{4}{1 + \beta}) \ln 4 + 4\epsilon N^{-\gamma}}
\]
\[
\leq 5N^{-\frac{1}{2} + \delta},
\]
where (a) follows from \( 1 - \frac{\epsilon N^{-\gamma}}{1-N^{-\gamma}} = \frac{\epsilon N^{-\gamma}}{1-N^{-\gamma}} \) and (b) follows from \( N^{-\gamma} \leq 2^{-1/2} \leq \frac{1}{2} \) and choosing \( q \geq q_0 \) with \( \ln q_0 = \frac{1}{q}(2\beta \ln 4 - \ln \beta + 2 \ln 6 - \ln 4) \). This completes the proof.

A. Numerical Examples

In this section, we present some applications of Corollary 5 based on numerical computation of \( \lambda \).

Example 8. Consider the case of \( q = 2 \) where \( T_q \) is defined by
\[
(T_2 g)(x) = \frac{g(x^2) + g(2x-x^2)}{2}.
\]
Using $V(x) = (x(1-x))^{0.66}$, one can verify numerically that $(T_2V)(x) \leq 0.832V(x)$ for $x \in [0,1]$. For example, see Figure 1. Therefore,

$$P\left(X_n \in [0.01, 0.99] \bigg| X_0 = x \right) \leq \left(\frac{1/4}{0.0099}\right)^{0.66} 0.832^n \leq 9 \cdot 2^{n \ln \frac{0.832}{\ln 2}} \leq 9N^{-0.265}.$$ 

Let $V_3(x)$ be the result of applying $T_2$ five times to the function $(x(1-x))^{0.66}$ (i.e., $V_3(x) = \left(T_2^5 (x(1-x))^{0.66}\right)(x)$). Then, one can verify numerically that $(T_2V_3)(x) \leq 0.8271V_5(x)$ and this gives a decay rate of $O(N^{-0.273})$.

**Example 9.** Consider the case of $q = 4$ where $T_q$ is defined by

$$(T_qg)(x) = \frac{1}{4} \left( g(x^4) + g(4x^3(1-x) + x^4) + g(1-4x(1-x)^3 - (1-x)^4) + g(1-(1-x)^4) \right).$$

Using $V(x) = (x(1-x))^{0.64}$, one can verify numerically that $(T_qV)(x) \leq 0.657V(x)$ for $x \in [0,1]$. Therefore,

$$P\left(X_n \in [0.01, 0.99] \bigg| X_0 = x \right) \leq \left(\frac{1/4}{0.0099}\right)^{0.64} 0.657^n \leq 8N^{-0.303}.$$ 

**Example 10.** Consider the case of $q = 16$ where $T_q$ is defined by (8). Using $V(x) = (x(1-x))^{0.58}$, one can verify numerically that $(T_qV)(x) \leq 0.375V(x)$ for $x \in [0,1]$. For example, see Figure 2. Therefore,

$$P\left(X_n \in [0.01, 0.99] \bigg| X_0 = x \right) \leq \left(\frac{1/4}{0.0099}\right)^{0.58} 0.375^n \leq 7 \cdot 16^{n \ln \frac{0.375}{\ln 2}} \leq 7N^{-0.353}.$$ 

**Example 11.** Consider the case where $T_q$ is defined by (2) for $q = 2, 3, \ldots, 1024$. Using $V(x) = \sqrt{x(1-x)}$, one can compute numerically the smallest $\lambda_q$ such that $(T_qV)(x) \leq \lambda_qV(x)$ for $x \in [0,1]$. This computation results in $\lambda_q = (T_qV)(\frac{1}{2})/V(\frac{1}{2})$ and one observes that $\sqrt{q}\lambda_q$ is increasing in $q$ and upper bounded by $1.6142$. Assuming this is true, we observe that

$$P\left(X_n \in [\eta, 1-\eta] \bigg| X_0 = x \right) \leq \frac{\sqrt{1/4}}{\sqrt{\eta(1-\eta)}} \left(\frac{1.6142}{\sqrt{q}}\right)^n \leq \frac{1}{\sqrt{4\eta(1-\eta)}} N^{-\frac{1}{2}(1 - \frac{1}{4\eta})}.$$ 

**Example 12.** Let $V(x) = (x(1-x))^{1/12}$ and consider the case where $T_q$ is defined by (3) for $q = 2, 3, \ldots, 1024$. Again, one can compute numerically the smallest $\lambda_q = \lambda_{q,1/12}$ such that $(T_qV)(x) \leq \lambda_qV(x)$ for $x \in [0,1]$. This computation results in $\lambda_q = (T_qV)(\frac{1}{2})/V(\frac{1}{2})$ and one observes that $\sqrt{q}\lambda_q$ is increasing in $q$ and upper bounded by $4.1218$. Assuming this is true, we observe that, for $N \geq 2$, we have

$$P\left(X_n \in [N^{-2}, 1-N^{-2}] \bigg| X_0 = x \right) \leq \frac{(1/4)^{1/12}}{(N^{-2}(1-N^{-2}))^{1/12}} \left(\frac{4.1218}{\sqrt{q}}\right)^n \leq \left(\frac{1/4}{1-N^{-2}}\right)^{1/12} N^{\frac{1}{2}} 6^{n \ln \frac{4.1218}{\ln q}} \leq \frac{1}{3} N^{-\frac{1}{2}(1 - \frac{1}{4\eta})}.$$ 

### III. LARGE ALPHABET ERASURE CHANNELS

#### A. Intuitive Approach

Before delving into the proof of Theorem 6, we present an intuitive (but non-rigorous) argument that leads us in the right direction. Consider $q$ random trials with success probability $x$ and let the random variable
Bin$(q, x)$ denote number of successes. Then, one finds that
\[
\mathbb{P}(\text{Bin}(q, x) = i) = \binom{q}{i} x^i(1 - x)^{q-i}.
\]

The key is to replace the binomial random variable, Bin$(q, x)$, by a Gaussian random variable with the same mean and variance. While this step is motivated by the central limit theorem, it is not rigorous (even as $q \to \infty$) because the approximation does not hold uniformly for all $x \in [0, 1]$. Based on this assumption, we approximate $\mathbb{P}(\text{Bin}(q, x) \geq i + 1)$ by
\[
\psi_i(x) \approx Q\left(\frac{i + 1 - qx}{\sqrt{qx(1-x)}}\right),
\]
where $Q(x) \triangleq (2\pi)^{-1/2} \int_{-}\infty^{\infty} e^{-t^2/2} dt$. Let $V(x) = (x(1-x))^\beta$ for $\beta \in (0, 2)$. Using a sequence of approximations, one finds that
\[
(T_q V)(x) = \frac{1}{q} \sum_{i=0}^{q-1} V(\psi_i(x))
\approx \frac{1}{q} \sum_{i=0}^{q-1} V\left(Q\left(\frac{i + 1 - qx}{\sqrt{qx(1-x)}}\right)\right)
\approx \int_0^1 V\left(Q\left(\frac{q(y-x)}{\sqrt{qx(1-x)}}\right)\right) dy
= \sqrt{\frac{q}{x(1-x)}} \int_{-}\infty^{\infty} V(Q(z)) dz
\approx \sqrt{\frac{q}{x(1-x)}} \int_{-}\infty^{\infty} V(Q(z)) dz
= \sqrt{\frac{q}{x(1-x)}} m(\beta),
\]
where $m(\beta) \triangleq \int_{-}\infty^{\infty} (Q(z)Q(-z))^{\beta} dz$.

For $\beta = \frac{1}{2}$, this implies that $V(x) = \sqrt{x(1-x)}$ is an approximate eigenfunction of $T_q$ associated with eigenvalue
\[
\hat{\lambda}_q = \frac{m(\frac{1}{2})}{\sqrt{q}},
\]
where $m(\frac{1}{2}) \approx 1.6147 < e^{1/2}$. Based on this estimate, one could estimate that rate of polarization scales like
\[
\hat{\lambda}_q^n \approx \left(\frac{m(\frac{1}{2})}{\sqrt{q}}\right)^n \leq \left(\frac{e}{q}\right)^{n/2}
= q^{(n/2)(1-\ln q)/\ln q} = N^{-\frac{1}{2}(1-\frac{1}{\ln q})}.
\]

In fact, the numerical results in Example [1] support this conclusion and suggest that the true $\lambda_{q,1/2} \leq m(\frac{1}{2})$. Thus, we believe that this non-rigorous analysis produces an exact and tight characterization as $q \to \infty$.

For $\beta \in (0, \frac{1}{2})$, $V(x)$ is not an approximate eigenfunction but one can still estimate the decay rate
\[
\hat{\lambda}_{q,\beta} = \max_{x \in [0,1]} (x(1-x))^{\frac{1}{2} - \beta} \frac{m(\beta)}{\sqrt{q}}
\leq \frac{m(\beta)}{\sqrt{q}} \left(\frac{1}{4}\right)^{\frac{1}{2} - \beta}.
\]

Combining this estimate with Corollary [5] gives the non-rigorous prediction
\[
\mathbb{P}(X_n \in [N^{-\gamma}, 1 - N^{-\gamma}]) | X_0 = x)
\leq \frac{(x(1-x))^{\beta}}{(1 - N^{-\gamma})^{\beta}} \frac{m(\beta)}{\sqrt{q}} \left(\frac{1}{4}\right)^{\frac{1}{2} - \beta}.
\]

This implies the following lemma.

**Lemma 13.** If $g(x) = g(1-x)$, then $(T_qg)(x) = (T_qg)(1-x)$.

**Proof:** Working directly, one finds that
\[
(T_qg)(x) = \frac{1}{q} \sum_{i=0}^{q-1} g(\psi_i(x))
= \frac{1}{q} \sum_{i=0}^{q-1} g(1 - \psi_i(x))
= \frac{1}{q} \sum_{i=0}^{q-1} g(\psi_{q-i-1}(1-x)).
\]
\[
\psi_i(x) = \begin{cases} 
\sum_{i=0}^{q-1} g(\psi_i(1-x)) \\
(T_q g)(1-x).
\end{cases}
\]

The well-known Chernoff bound for the binomial tail probability implies that, for \(i + 1 \geq qx\), one has
\[
\psi_i(x) = P(Bin(q, x) \geq i + 1) \leq e^{-qD(\frac{1}{q}||x)}, \tag{4}
\]
where \(D(y||x) \triangleq y \ln \frac{y}{x} + (1-y) \ln \frac{1-y}{1-x}\) is the Kullback-Leibler divergence between two Bernoulli distributions. Similarly, for \(i \leq qx\), one has
\[
1 - \psi_i(x) = P(Bin(q, x) \leq i) \leq e^{-qD(\frac{1}{q}||x)}. \tag{5}
\]

**Lemma 14.** For \(x \leq y\), we have
\[
P(Bin(q, x) \geq qy) \leq e^{-qD(y,x)},
\]
where \(D(y,x) \triangleq \frac{1}{2}(y-x)^2/(x(1-x) + (1-2x)(y-x)/3)\). Similarly, for \(x \geq y\), we have \(P(Bin(q, x) \leq qy) \leq e^{-qD(y,x)}\).

**Proof:** It is well known from the Chernoff bound that \(P(Bin(q, x) \geq qy) \leq e^{-qD(y,x)}\) for \(x \leq y\), where \(D(y||x) \triangleq y \ln \frac{y}{x} + (1-y) \ln \frac{1-y}{1-x}\) is the Kullback-Leibler divergence. Thus, the first result holds if \(d(y,x) \leq D(y||x)\) for \(x \leq y\). Since \(D(y||1-x) = D(y||x)\) and \(d(1-y,1-x) = d(y,x)\), the second result follows from the first by symmetry. Thus, it suffices to prove that \(d(y,x) \leq D(y||x)\) for \(x \leq y\). To do this, we first observe that
\[
\frac{d}{dx} (d(y,x) - D(y||x)) = \frac{(1-x)(1-x)/2}{x(1-x)(y-x(x+2y-2))} \geq 0
\]
for \(x \leq y\). Next, we observe that
\[
\int_x^y \left( \frac{d}{dx} (d(y,x') - D(y||x')) \right) dx = (d(y,x) - D(y||y)) - (d(y,x) - D(y||x)) \geq 0
\]
because \(x \leq y\) throughout the range of integration. Since \(d(y,y) = D(y||y) = 0\), this implies \(d(y,x) \leq D(y||x)\) for \(x \leq y\).

**Lemma 15.** For \(x, y \in [0, 1]\), we have
\[
D(y||x) \geq (y-x) + (1-y) \ln \frac{1-y}{1-x}
\]
and, for \(z \in [0, 1]\), we have
\[
1 - z + z \ln z \geq \frac{1}{2}(1-z)^2.
\]

**Proof:** The first bound follows from lower bounding the \(y \ln \frac{y}{x}\) term in \(D(y||x)\) by
\[
y \ln \frac{y}{x} = -y \ln \frac{y-x}{y} \geq -y \left( \frac{y-x}{y} \right) = -x.
\]

Let \(f(z) = z + (1-z) \ln (1-z)\) and observe that
\[
f(1-z) = 1 - z + z \ln z.
\]
Since \(f'(0) = f(0) = 0\) and \(f''(z) = \frac{1}{1-z} \geq 1\) for \(z \in [0, 1]\), it follows that
\[
f(z) = \int_0^z \int_0^y f''(x) dx \ dy \geq \frac{1}{2}z^2.
\]
Thus, \(f(1-z) = 1 - z + z \ln z \geq \frac{1}{2}(1-z)^2\).

**Lemma 16.** For \(\beta \in (0, \frac{1}{2}]\), \(V(x) = (x(1-x))^\beta\), and \(x \in [\frac{1}{2}, 1]\), we have
\[
\frac{1}{q} V(\psi_{[qx]} - 1(x)) \leq \frac{2(x(1-x))^\beta}{\sqrt{2q}}.
\]

**Proof:** If \(x \in [\frac{1}{2}, 1 - \frac{1}{q}]\), then we have
\[
\frac{1}{q} V(\psi_{[qx]} - 1(x)) \leq \frac{2(x(1-x))^\beta}{\sqrt{2q}} \leq \frac{2(x(1-x))^\beta}{\sqrt{2q}} \frac{1}{q} \left( \frac{1}{4} \right)^\beta
\]
\[
\leq \frac{2(x(1-x))^\beta}{\sqrt{2q}} \frac{1}{q} \left( \frac{1}{4} \right)^\beta
\]
\[
\leq \frac{2(x(1-x))^\beta}{\sqrt{2q}} \sup_{q \geq 2} q^{1/2} 4^\beta
\]
\[
= \frac{(x(1-x))^\beta}{\sqrt{2q}}.
\]
where \((a)\) holds because \(V(z) \leq \left( \frac{1}{4} \right)^\beta\), \((b)\) follows from \(2x \geq 1\) and \(q(1-x) \geq 1\), and \((c)\) holds because the argument of the supremum is decreasing in \(q\). If \(x \in (1 - \frac{1}{2}, 1]\), then assume \(x = 1 - \frac{\alpha}{q}\) for \(\alpha \in [0, 1]\) and observe that
\[
\frac{1}{q} V(\psi_{[qx]} - 1(x)) = \frac{1}{q} V(\psi_{[q-1]}(x)) = \frac{1}{q} V(x(1-x)^\beta)
\]
\[
= \frac{1}{q} \left( \frac{1}{q} (1-x)^\beta \right)^\beta
\]
\[
= \frac{1}{q} \left( \frac{1}{q} \left( \frac{1}{q} \right)^q \left( 1 - \left( \frac{1}{q} \right)^q \right) \right)^\beta
\]
\[
= \frac{1}{q} e^{-q \alpha} \alpha
\]
\[
= \frac{\left( \frac{1}{q} \left( 1 - \alpha \right) \right)^\beta}{\sqrt{q}} \frac{1}{q} e^{-q \alpha} \alpha
\]
\[
= \frac{\left( \frac{1}{q} \left( 1 - \alpha \right) \right)^\beta}{\sqrt{q}} \frac{q^{1/2}}{\left( \frac{1}{q} \right)^\beta} e^{-q \alpha}
\]
\[
\leq \frac{(x(1-x))^\beta}{\sqrt{2q}} \sup_{q \geq 2} q^{1/2} 4^\beta e^{-q \alpha}
\]
\[
= \frac{(x(1-x))^\beta}{\sqrt{2q}}.
\]
\( \frac{(x(1-x))^\beta}{\sqrt{q}} \sup_{\alpha \in [0,1]} \left( (1 - \frac{q}{2}) \right)^\beta e^{-\alpha \beta} \leq \frac{(x(1-x))^\beta}{\sqrt{q}} 2^{\beta - 1/2}, \)

where (a) follows from \( 1 - \alpha \leq \left( 1 - \frac{q}{2} \right)^q \leq e^{-\alpha} \), (b) holds because the argument of the supremum is decreasing in \( q \), and (c) holds because the argument of the supremum is decreasing in \( \alpha \).

C. Proof of Lemma 5

Let \( V(x) = (x(1-x))^\beta \) with \( \beta \in (0, \frac{1}{2}] \). Based on Lemma 13, it is sufficient to analyze \( (T_q V)(x) \) for \( x \geq 1/2 \). To do this, we will use the decomposition

\[
\frac{1}{q} \sum_{i=0}^{[qx]-2} V(q(i+1)x) = \frac{1}{q} \left( \sum_{i=0}^{[qx]-2} V(q(i)x) \right) + V(q([qx]-1)x) + \sum_{i=1}^{[q]-1} V(q(i)x). \tag{6}
\]

First, we consider the upper sum in (6). Applying (4) to \( \psi_i(x) \) shows that

\[
\psi_i(x) = \mathbb{P}(\text{Bin}(q, x) \geq i + 1) \leq e^{-qD(\frac{i+1}{q} || x)}
\]

for \( i + 1 \geq qx \). Thus, for \( i \in \{[qx], \ldots, [q]-1\} \), we have \( V(\psi_i(x)) \leq (\psi_i(x))^{\beta} \leq e^{-qD(\frac{i+1}{q} || x}) \) and

\[
\frac{1}{q} \sum_{i=0}^{[q]-2} V(q(i)x) \leq \frac{1}{q} \sum_{i=0}^{q-1} e^{-qD(\frac{i+1}{q} || x}) \leq \frac{1}{q} \sum_{i=0}^{q-1} e^{-qD(\frac{i+1}{q} \frac{q}{q} || x}) d\gamma \]

\[
\leq \frac{1}{q} \int_{[q]-1/q}^{1} e^{-qD(\frac{i+1}{q} || x}) d\gamma \]

where \( e^{-qD(\frac{i+1}{q} || x}) \leq \int_0^1 e^{-qD(\frac{i+1}{q} || x}) d\gamma \) holds in (a) because \( e^{-qD(\frac{i+1}{q} || x}) \) is decreasing in \( z \) for \( i \geq qx \). Also, (b) follows from grouping terms into one integral and changing the variable of integration. Although this bound holds for all \( x \in [0,1] \), the sum is empty for \( x \in (1 - \frac{1}{q}, 1] \) and trivially equal to zero.

For \( x \geq \frac{1}{2} \), an upper bound on the integral is given by

\[
\int_{[q]-1/q}^{1} e^{-qD(\frac{i+1}{q} || x}) d\gamma \leq \int_{x}^{x} e^{-qD(\frac{i+1}{q} || x}) d\gamma \leq \int_{x}^{x} \exp \left( \frac{-q\beta(y-x^2)}{2(x(1-x) + (2x-1)(x-y)/3) \beta} \right) dy \leq \int_{2x-1}^{x} \exp \left( \frac{-q\beta(y-x^2)}{2((1-x) + (1-x)/3) \beta} \right) dy
\]

where (a) follows from Lemma 14 and (b) holds because \( (1 - 2x)(y - x) \leq 0 \) for \( y \geq x \).

Now, we consider the lower sum in (6). Similarly, for \( i \in \{0, \ldots, [qx] - 2\} \), (5) shows that

\[
\psi_i(x) = 1 - \mathbb{P}(\text{Bin}(q, x) \leq i) \geq 1 - e^{-qD(\frac{i}{q} || x}).
\]

For \( i \in \{0, \ldots, [qx] - 2\} \), we have \( V(1 - \psi_i(x)) \leq (1 - \psi_i(x))^{\beta} \leq e^{-qD(\frac{i}{q} || x}) \) and thus

\[
\frac{1}{q} \sum_{i=0}^{[q]-2} V(q(i)x) = \frac{1}{q} \sum_{i=0}^{[q]-2} V(1 - \psi_i(x)) \leq \frac{1}{q} \sum_{i=0}^{[q]-2} e^{-qD(\frac{i}{q} || x}) \leq \frac{1}{q} \sum_{i=0}^{[q]-2} \int_0^1 e^{-qD(\frac{i}{q} || x}) dx \]

where \( e^{-qD(\frac{i}{q} || x}) \leq \int_0^1 e^{-qD(\frac{i+1}{q} || x}) d\gamma \) holds in (a) because \( e^{-qD(\frac{i+1}{q} || x}) \) is increasing in \( z \) for \( z \in [0,1] \) and \( i + 1 \leq qx \). Also, (b) follows from grouping terms into one integral and changing the variable of integration.

The expression in (9) can be upper bounded using the decomposition

\[
\int_0^{([q]-1)/q} e^{-qD(y || x}) dy \leq \int_{2x-1}^{x} e^{-qD(y || x}) dy + \int_{2x-1}^{x} e^{-qD(y || x}) dy
\]

where \( e^{-qD(\frac{i+1}{q} || x}) \leq \int_0^1 e^{-qD(\frac{i+1}{q} || x}) d\gamma \) holds in (a) because \( e^{-qD(\frac{i+1}{q} || x}) \) is decreasing in \( z \) for \( i \geq qx \). Also, (b) follows from grouping terms into one integral and changing the variable of integration.

The first term in (10) can be upper bounded with

\[
\int_{2x-1}^{x} e^{-qD(y || x}) dy \leq \int_{2x-1}^{x} \exp \left( \frac{-q\beta(y-x^2)}{2(x(1-x) + (2x-1)(x-y)/3) \beta} \right) dy \leq \int_{2x-1}^{x} \exp \left( \frac{-q\beta(y-x^2)}{2((1-x) + (1-x)/3) \beta} \right) dy
\]
\[ \sqrt{\frac{2\pi(1-x)}{3q\beta}} \text{erf} \left( \sqrt{\frac{3\beta q(1-x)}{8}} \right) \]
\[ \leq \sqrt{\frac{4\pi x(1-x)}{3q\beta}}. \quad (11) \]

where (a) follows from Lemma 14, (b) holds because \( x(1-x) \leq 1-x \) and \( (2x-1)(x-y) \leq 1-x \) for \( y \geq 2x-1 \) and \( x \geq \frac{1}{2} \), and (c) follows from \( 2x \geq 1 \) for \( x \geq \frac{1}{2} \). The second term in (10) can be upper bounded with

\[ \int_0^{2x-1} e^{-\beta D(y|x)} dy \]
\[ \leq \int_0^{2x-1} \exp \left( -\beta \left( (y-x) + (1-y) \ln \frac{1-y}{1-x} \right) \right) dy \]
\[ = (1-x) \int_2^{1-x} \exp \left( \frac{1}{2} \beta (x-1) (z-1)^2 \right) dz \]
\[ \leq \pi \sqrt{\frac{2q\beta(1-x)}{2}} \text{erf} \left( \frac{q\beta(1-x)(z-1)^2}{2} \right) \mid_{z=2}^{1-x} \]
\[ \leq \sqrt{\frac{\pi x(1-x)}{q\beta}}, \quad (12) \]

where (a) follows from Lemma 15, (b) is given by the change of variables \( y \mapsto 1-z(1-x) \), (c) holds because \( 1-z + z \ln z \geq \frac{3}{2}(z-1)^2 \) for \( z \in [0, 1] \), (d) follows from \( \text{erf}(b) - \text{erf}(a) \leq 1 \) for \( b \geq a \geq 0 \), and (e) holds because \( 2x \geq 1 \) for \( x \geq \frac{1}{2} \).

Now, we combine Lemma 16 with (8), (11), and (12) to see that

\[ (T_q V)(x) = \frac{1}{q} \left[ \sum_{i=0}^{[q\beta]-2} V(\psi_i(x)) + \sum_{i=[q\beta]}^{q-1} V(\psi_i(x)) \right] \]
\[ \leq \int_0^{([q\beta]-1)/q} e^{-\beta D(y|x)} dy + \frac{1}{q} V(\psi_{[q\beta]-1}(x)) \]
\[ + \int_{[q\beta]/q}^{1} e^{-\beta D(y|x)} dy \]
\[ \leq \sqrt{\frac{\pi x(1-x)}{q\beta}} + \sqrt{\frac{4\pi x(1-x)}{3q\beta}} + \frac{(2x(1-x))^\beta}{\sqrt{2q}} + \sqrt{\frac{\pi x(1-x)}{2q\beta}} \]
\[ \leq (2x(1-x))^\beta + A \sqrt{x(1-x)} \]

where \( A = \sqrt{\pi+\sqrt{\frac{4\pi}{3}}} + \sqrt{\frac{\pi}{2}} \). Combining Definition 3 with (13), we see that

\[ \lambda_{q,\beta} \eqdef \sup_{x \in (0,1)} \frac{(T_q V)(x)}{V(x)} \]
\[ \leq \sup_{x \in (0,1)} \left( \frac{2^\beta}{\sqrt{2q}} + \frac{A}{\sqrt{q\beta}} (x(1-x))^{1/2-\beta} \right) \]
\[ \leq \frac{2^\beta}{\sqrt{2q}} + A \left( 1 + \frac{2^\beta \sqrt{3}}{\sqrt{2} (\frac{1}{4})^{1/4}} \right) \]
\[ \leq \frac{6}{\sqrt{q\beta}} \left( \frac{1}{4} \right)^{3/4} \leq 6 \]

where (a) holds because the supremum is achieved at \( x = \frac{1}{2} \) and (b) holds because

\[ A + \frac{2^\beta \sqrt{3}}{\sqrt{2} (\frac{1}{4})^{3/4}} \leq 6 \]

for \( \beta \in (0, \frac{1}{2}) \).

**IV. Conclusion**

In this paper, we investigate the relationship between the blocklength and the gap to capacity for the \( q \)-ary Reed-Solomon polar codes introduced by Mori and Tanaka. These codes have length \( N = q^n \), where \( n \) is the number of steps in the polarization process. When one of these codes is transmitted over a \( q \)-ary erasure channel with erasure probability \( \epsilon \), its effective channels are \( q \)-ary erasure channels and their erasure rate satisfy a closed-form recursion. By analyzing this recursion, we show that, for any \( \gamma > 0 \) and \( \delta > 0 \), there is a \( q_0 \) such that, for all \( q \geq q_0 \), the fraction of effective channels with erasure rate at most \( O(N^{-\gamma}) \) is at least \( 1 - \epsilon - O(N^{-1/2+\delta}) \). Thus, the gap to capacity scales at a rate very close to the optimal rate of \( O(N^{-1/2}) \).

This naturally suggests two interesting open questions. First, can this result be extended to noisy \( q \)-ary channels? Second, can one prove that binary polar codes with \( \ell \times \ell \) polarization kernels also achieve near-optimal scaling on the BEC as \( \ell \to \infty \)? This question is discussed in some detail in [21].
REFERENCES

[1] V. Strassen, “Meßfehler und information,” Zeitschrift für Wahrscheinlichkeitstheorie, vol. 2, no. 4, pp. 273–305, 1964.
[2] J. N. Laneman, “On the distribution of mutual information,” in Proc. 1st Annual Workshop on Inform. Theory and its Appl., (San Diego, CA, USA), Feb. 2006.
[3] M. Hayashi, “Information spectrum approach to second-order coding rate in channel coding,” IEEE Trans. Inform. Theory, vol. 55, no. 11, pp. 4947–4966, 2009.
[4] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” IEEE Trans. Inform. Theory, vol. 56, no. 5, pp. 2307–2359, 2010.
[5] E. Arikan, “Channel polarization: A method for constructing capacity-achieving codes,” in Proc. IEEE Int. Symp. Inform. Theory, (Toronto, Canada), pp. 1173–1177, July 2008.
[6] E. Arikan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” IEEE Trans. Inform. Theory, vol. 55, pp. 3051–3073, July 2009.
[7] E. Arikan and E. Telatar, “On the rate of channel polarization,” in Proc. IEEE Int. Symp. Inform. Theory, pp. 1493–1495, 2009.
[8] S. H. Hassani, R. Mori, T. Tanaka, and R. L. Urbanke, “Rate-dependent analysis of the asymptotic behavior of channel polarization,” IEEE Trans. Inform. Theory, vol. 59, no. 4, pp. 2267–2276, 2013.
[9] D. Goldin and D. Burshtein, “Improved bounds on the finite length scaling of polar codes,” IEEE Trans. Inform. Theory, vol. 60, no. 11, pp. 6966–6978, 2014.
[10] S. H. Hassani, K. Alishahi, and R. Urbanke, “Finite-length scaling for polar codes,” IEEE Trans. Inform. Theory, vol. 60, no. 10, pp. 5875–5898, 2014.
[11] A. Fazeli and A. Vardy, “On the scaling exponent of binary polarization kernels,” in Proc. Annual Allerton Conf. on Commun., Control, and Comp., pp. 797–804, 2014.
[12] V. Guruswami and A. Velingker, “An entropy subset inequality and polynomially fast convergence to Shannon capacity over all alphabets,” arXiv preprint arXiv:1411.6993 2014.
[13] N. Presman, O. Shapira, S. Litsyn, T. Etzion, and A. Vardy, “Binary polarization kernels from code decompositions,” IEEE Trans. Inform. Theory, vol. 61, no. 5, pp. 2227–2239, 2015.
[14] V. Guruswami and P. Xia, “Polar codes: Speed of polarization and polynomial gap to capacity,” IEEE Trans. Inform. Theory, vol. 61, no. 1, pp. 3–16, 2015.
[15] D. Goldin and D. Burshtein, “On the finite length scaling of ternary polar codes.” arXiv preprint arXiv:1502.02925 2015.
[16] M. Mondelli, S. H. Hassani, and R. Urbanke, “Unified scaling of polar codes: Error exponent, scaling exponent, moderate deviations, and error floors.” [Online]. Available: http://arxiv.org/abs/1501.0244 2015.
[17] R. Mori and T. Tanaka, “Non-binary polar codes using Reed-Solomon codes and algebraic geometry codes,” in Proc. IEEE Inform. Theory Workshop, pp. 1–5, Aug. 2010.
[18] R. Mori and T. Tanaka, “Source and channel polarization over finite fields and Reed-Solomon matrices,” IEEE Trans. Inform. Theory, vol. 60, no. 5, pp. 2720–2736, 2014.
[19] F. R. Kschischang, “Constructing Reed-Muller codes from Reed-Solomon codes over GF(q),” in Proc. IEEE Int. Symp. Inform. Theory, pp. 195–195, 1993.
[20] M. Hairer and J. C. Mattingly, “Yet another look at Harris’ ergodic theorem for Markov chains,” in Seminar on Stochastic Analysis, Random Fields and Applications VI, pp. 109–117, Springer, 2011.
[21] S. H. Hassani, Polarization and spatial coupling: Two techniques to boost performance. PhD thesis, École Polytechnique Fédérale de Lausanne, 2013.