Concatenated Polar Codes

Mayank Bakshi[1], Sidharth Jaggi[2], Michelle Effros[1]
[1] California Institute of Technology [2] Chinese University of Hong Kong

***Abstract*—Polar codes have attracted much recent attention as one of the first codes with low computational complexity that provably achieve optimal rate-regions for a large class of information-theoretic problems. One significant drawback, however, is that for current constructions the probability of error decays sub-exponentially in the block-length (more detailed designs improve the probability of error at the cost of significantly increased computational complexity. In this work we show how the the classical idea of code concatenation – using "short" polar codes as inner codes and a "high-rate" Reed-Solomon code as the outer code – results in substantially improved performance. In particular, code concatenation with a careful choice of parameters boosts the rate of decay of the probability of error to almost exponential in the block-length with essentially no loss in computational complexity. We demonstrate such performance improvements for three sets of information-theoretic problems – a classical point-to-point channel coding problem, a class of multiple-input multiple output channel coding problems, and some network source coding problems.**

I. Introduction

Polar codes [1] are provably capacity-achieving codes for the Binary Symmetric Channel, with code complexities that scale as $\mathcal{O}(N \log N)$ in the block-length N. Polar codes have since demonstrated their versatility. Capacity-achieving low-complexity schemes based on polar codes have been demonstrated for a wide variety of source and channel coding problems. Examples include some point-to-point discrete memoryless channels, some rate-distortion problems, the Wyner-Ziv problem and the Gelfand-Pinsker problem [2].

A significant drawback remains. The minimum distance of the polar codes in [1] is shown in [3] to grow no faster than $o(\sqrt{N})$ in the block-length N. This is used to show [3] that the probability of error decays no faster than $\exp(-o(\sqrt{N}))$ (compared with the $\exp(-\Theta(N))$ probability of error provable for random codes [4]). Low-complexity codes achieving this performance have been constructed [5]. Further work to improve the decay-rate of the error probability was partially successful – a sequence of codes have been constructed [6] that in the limit (of a certain implicit parameter denoted l) achieves $\exp(-o(N))$ probability of error; however this improvement comes at the expense of significantly increased computational complexity, which scales as $\mathcal{O}(2^l N \log N)$.

In this work we demonstrate that concatenating short polar codes with a high-rate outer Reed-Solomon code significantly improves the rate of decay of the probability of error, with little or no cost in computational complexity. The price we pay is that the rate of convergence with block-length N of our codes to the information-theoretically optimal rates is slower than that of polar codes. For the point-to-point channel coding problem we use capacity-achieving polar codes of block-length $\Theta(\log^3 N)$ as the inner codes. The overall encoding procedure is linear over the binary field.

There are three cases of interest. The first case is at one extreme, in which "many" of the inner codes (at least a $\log^{-3/2} N$ fraction) fail, resulting in a decoding error with probability $\exp(-\Omega(N/(\log^{-27/8} N)))$. This is the only scenario in which our scheme decodes erroneously.

The second scenario is at the other extreme, in which *none* of the inner codes fail. We show here that if the outer code is a systematic Reed-Solomon code with a rate that approaches 1 asymptotically as a function of N, the decoder can quickly verify and decode to the correct output with computational complexity $\mathcal{O}(N(poly(\log N)))$. We show that this is the likeliest scenario since it occurs with probability $1 - o(1/N)$.

The third scenario is the intermediate regime in which at least one (but fewer than a $\log^{-3/2} N$ fraction) of the inner codes fail. Here we show that the Reed-Solomon outer code can correct the errors in the outputs of the inner codes. The complexity of decoding in this scenario is $\mathcal{O}(N^2)$, which is dominated by the Berlekamp-Massey decoding algorithm for Reed-Solomon codes [7]. However, since this scenario occurs with probability $o(1/N)$, the *average* computational complexity is still dominated by the second scenario.

We then extend these techniques to two other classes of problems. The first class is a general class of multiple-input multiple-output channels, which include as special cases the multiple-access channel and the degraded broadcast channel. The second class is that of network source coding, which includes as a special case the Slepian-Wolf problem [8]. Prior to polar codes, no provable low-complexity capacity achieving schemes were known that achieved the optimal rates for these problems. In all cases our codes improve on the rate of decay of the probability of error that polar codes attain, while leaving other parameters of interest essentially unchanged.

Our concatenated code constructions preserve the linearity of the polar codes they are built upon. This is because both the inner polar codes and the outer Reed-Solomon code have a linear encoding structure, albeit over different fields ($\mathbb{F}_2$ and $\mathbb{F}_q$ respectively). However, if we choose the field for the outer code so that $q = 2^r$ for some integer r, all linear operations required for encoding over $\mathbb{F}_q$ may be implemented as $r \times r$ matrix operations over $\mathbb{F}_2$. Hence the encoding procedures for both the inner and the outer code may be composed to form

This work was supported in part by RGC GRF grants 412608 and 412809, RGC AoE grant on Institute of Network Coding, established under the University Grant Committee of Hong Kong, CUHK MoE-Microsoft Key Laboratory of Humancentric Computing and Interface Technologies, DARPA ITMANET program under grant W911NF-07-1-0029, and Caltech's Lee Center for Advanced Networking.

a code that is overall a linear code over $\mathbb{F}_2$.

II. Background

A. Polar codes [1] are provably capacity-achieving codes with low encoding and decoding complexity for arbitrary binary-input symmetric discrete memoryless channels. To simplify presentation we focus on binary symmetric channels [8] (BSC(p)) though many of the results can be generalized to other channels [2]. A crucial component of a polar code is a binary $l \times l$ "base matrix" denoted G which defines many of their properties. We replicate here some such important properties relevant for this work.

Since polar codes are used as inner codes in our construction, we call its rate the *inner code rate*, which we denote by R_I. The polar encoder takes as input $R_I n$ bits, and outputs n bits into the channel. The BSC(p) channel flips each bit independently with a probability p. The polar decoder then attempts to reconstruct the encoder's $R_I n$ bits.

Probability of error: The best known rate of decay of the probability of error of polar codes with increasing block-length N (see [2], [5], [6]) is $(\exp(-o(N^{\beta(l)})))$. Here $\beta(l)$, called the *exponent of the polar code*, is challenging to compute,[1] but for instance it is known that $\beta(2) = 0.5$, $\beta(l) \leq 0.5$ for $l \leq 15$, and $\beta(l) \leq 0.6$ for $l \leq 30$.

Complexity: The encoding and the decoding complexities of polar codes are $\mathcal{O}(lN \log N)$ and $\mathcal{O}(2^l N \log N)$ respectively[2].

Rate: While the exact speed of convergence of the rate to the Shannon capacity is not known, it is known that polar codes are asymptotically rate-optimal. In this work we denote the (unknown) *rate redundancy* of polar codes by $\delta(n)$.

Other rate-optimal channel codes: There has been much attention on the excellent empirical performance (asymptotically capacity achieving, low encoding and decoding complexity, fast decay of the probability of error) of LDPC codes [9]. However, as of now these results are still not backed up by theoretical guarantees. On the other side, the state of the art in provably good codes are those of Spielman et al. [10], which are asymptotically optimal codes that have provably good performance in terms of computational complexity and probability of error. However, these codes too have their limitations – their computational complexity blows up as the rate of the code approaches capacity.

B. Reed-Solomon codes [11] (henceforth RS codes) are classical error-correcting codes. Let the *outer code rate* R_O of the RS code be any number in $(0, 1)$. The RS encoder takes as input $R_O m$ symbols[3] over a finite field $\mathbb{F}_q$ (here the rate R_O, the field-size q, and the outer code's block-length m are code design parameters to be specified later). The RS encoder outputs a sequence of m symbols over $\mathbb{F}_q$ that are then transmitted over the channel. The encoding complexity of RS-codes is low – clever implementations are equivalent to performing a *Fast Fourier Transform over* $\mathbb{F}_q$ [12]. This can be done with $\mathcal{O}(m \log m)$ operations over $\mathbb{F}_q$, or equivalently $\mathcal{O}(m \log m \log q)$ binary operations (for large q).

The channel is allowed to arbitrarily corrupt up to $m(1 - R_O)/2$ symbols. Given such a channel, for all $q \geq m$ the standard RS decoder [7] reconstructs the source message exactly. The fastest known RS decoders for such channels have time complexity $\mathcal{O}(m^2 \log m \log q)$ (for large q) [7]. In this work, we are interested in *systematic* RS codes [7]. In a systematic RS code the first $R_O m$ symbols are the same as the input to the RS encoder, and the remaining $(1 - R_O)m$ *parity-check* symbols correspond to the output of a generic RS encoder. These are used in our concatenated code constructions to give efficient decoding of "high-rate" RS codes when, with high probability no errors occur.

C. Code concatenation [13]–[15] proposed by Forney in 1966, means performing both encoding and decoding in two layers. The source information is first broken up into "many" chunks each of "small" size, and some redundancy is added via an *outer code*. Then each chunk is encoded via a separate *inner code*. The small block-length of each chunk results in a relatively high probability of error of the inner code but low code complexity. In contrast, since the outer code is over a large alphabet, it requires that error be concentrated in a "few" symbols for good performance. Given such conditions, it achieves a relatively small probability of error with efficient encoding and decoding algorithms (such as for RS codes). Combining the two layers with the appropriate choice of parameters results in an overall code with low computational complexity and fast decay of probability of error.

III. Main results

For ease of exposition we first outline our main ideas for a point-to-point binary symmetric channel BSC(p). Let the input block-length of the channel code be N.

Theorem 1. *For each N large enough there exists a (concatenated polar) code with computational complexity $\mathcal{O}(N \log N)$ that achieves the capacity of BSC(p) asymptotically in N, with a probability of error of at most* $\exp(-\Theta(N/\log^{27/8} N))$.

This can be extended to more general scenarios.

Corollary 1 (Multiple Access Channel and Degraded Broadcast Channel)**.** *For the two user multiple access channel $p_{Y|X_1X_2}(\cdot|\cdot)$ and the degraded broadcast channel $p_{X_1X_2|Y}(\cdot|\cdot)$, there exist asymptotically rate-optimal codes constructed by concatenating a polar code of block length $\log^3 N$ with a RS code of block length $N \log^{-3} N$ that have an error probability that decays as $\exp(-\Omega(N \log^{-27/8} N))$ and can be encoded and decoded in $\mathcal{O}(N \log N)$ time.*

In Section V we then give corresponding constructions for network source coding problems. Let the input block-length of the source codes be M.

[1] Upper and lower bounds on the growth of $\beta(l)$ with l are known [6] – these bounds are again not in closed form and require significant computation.

[2] We note that the rate of decay of the probability of error can be traded off with the computational complexity of polar codes. However, due to the exponential dependence of the computational complexity on the parameter l, this cost may be significant for codes that have an exponent close to 1.

[3] As is standard, we assume here that m/R_O is an integer – if not, choosing a large enough m allows one to choose $R'_O \approx R_O$ resulting in codes with approximately the same behavior in all the parameters of interest.

Parameter	Meaning	Our parameter choice
n	Block-length of inner codes (over $\mathbb{F}_2$)	$\log^3 N$
R_I	Rate of inner codes	$1 - H_b(p) - \delta(\log^3 N)$
p_i	Probability of error of inner codes	$\exp(-\Omega(\log^{9/8} N))$
$q = 2^{R_I n}$	Field-size of outer code	$2^{R_I \log^3 N}$
m	Block-length of outer code (over $\mathbb{F}_q$)	$N \log^{-3} N$
R_O	Rate of outer code	$1 - 2\log^{-3/2} N$
$N = nm$	Block-length of overall code (over $\mathbb{F}_2$)	N
$M = R_I R_O N$	Number of source bits	$\left(1 - H_b(p) - \delta(\log^3 N)\right)\left(1 - 2\log^{-3/2} N\right) N$
P_e	Probability of error of overall code	$\exp(-\Omega(N \log^{-27/8} N))$

Fig. 1. Summary of notation for the binary symmetric channel

Corollary 2 (Network Source Codes)**.** *Let $\mathcal{N}$ be a network such that there exists a sequence of polar codes that are asymptotically rate-optimal for some corresponding source-coding problem. Then, there exists a sequence of asymptotically rate-optimal concatenated codes for which, the error probability decays as* $\exp(-\Omega(M \log^{-27/8} M))$*. Further, the code complexity is* $\mathcal{O}(M \log M)$.

IV. Channel coding

As is well-known, the optimal rate achievable asymptotically in the block-length for a BSC(p) equals $1-H_b(p)$, where $H_b(.)$ refers to the *binary entropy function*.

A. Binary symmetric channel

Figure 1 provides a summary of code parameters.

1) Encoder: Let $n = \log^3 N$ be the inner polar codes' block-length, $R_I = 1-H_b(p)-\delta(n) = 1-H_b(p)-\delta(\log^3 N)$ be their rate, and $p_i = \exp(-\Omega(\log^{3\beta} N))$ be their probability of error. (Here β is any value[4] in $(1/3, 1/2)$. To be concrete, say $\beta = 3/8$.) Let $f_I : \mathbb{F}_2^{R_I n} \to \mathbb{F}_2^n$ and $g_I : \mathbb{F}_2^n \to \mathbb{F}_2^{R_I n}$ denote respectively their encoders and decoders.

Correspondingly, let $m = N/n = N/\log^3 N$ be the block-length of the outer systematic RS code, $q = 2^{R_I n} = 2^{R_I \log^3 N}$ be the field-size[5] and $R_O = 1 - 2\log^{-3/2} N$ be its rate (so the code can correct up to a fraction $\log^{-3/2} N$ of symbol errors). Let $f_O : \mathbb{F}_q^{R_O m} \to \mathbb{F}_q^m$ and $g_O : \mathbb{F}_{2^n}^m \to \mathbb{F}_{2^n}^{R_O m}$ denote respectively the encoder and (Berlekamp-Massey) decoder for the outer systematic RS code.

Let $M = R_O R_I N$. Define the concatenated code through the encoder function $f : \mathbb{F}_2^M \to \mathbb{F}_2^N$ such that for each *source message* $\mathbf{u}^M \in \{0,1\}^M$, $f(\mathbf{u}^M) = (f_I(\mathbf{x}_1), f_I(\mathbf{x}_2), \ldots, f_I(\mathbf{x}_m))$, where for each i in $\{1, 2, \ldots, m\}$, $\mathbf{x}_i$ represents the ith symbol of the output of the outer systematic RS encoder $f_O(\mathbf{u}^M)$, viewed as a length-$R_I n$ bit vector.

As noted in the introduction, since the inner code is linear over $\mathbb{F}_2$, and the outer code is linear over a field whose size is a power of 2 (and as such may be implemented via matrix operations over $\mathbb{F}_2$) the overall encoding operation is linear.

[4]The upper bound arises due to the provable rate of decay of the probability of error of polar codes [2], and the lower bound arises due to a technical condition required for (1) to hold.

[5]For this choice of parameters $q = \omega(m)$, as required for RS codes.

2) Channel: The channel corrupts each transmitted inner code vector $f_I(\mathbf{x}_i)$ to $\mathbf{y}_i$ resulting in the output $\mathbf{y}^N$.

3) Decoder: Our decoder:

1) Decodes each successive n-bit vector $\mathbf{y}_i$, $i \in \{1, \ldots, m\}$ using the inner polar code decoder g_I to the length-$R_I n$ bit vectors $\hat{\mathbf{x}}_i = g_I(\mathbf{y}_i)$, $i \in \{1, \ldots, m\}$

2) Passes the first $R_O m$ outputs $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_{R_O m}$ of the inner code decoders through the systematic RS encoder f_O.

3) If $f_O(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_{R_O m}) = \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_m$, it declares $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_{R_O m}$ as the decoded message (denoted by $\bar{\mathbf{x}}^M$) and terminates.

4) Otherwise it passes $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_m$ through the outer decoder g_O (a standard RS Berlekamp-Massey decoder), declares the length-M bit-vector corresponding to the output $\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_{R_O m}^* = g_O(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_m)$ as the decoded message (denoted by $\bar{\mathbf{x}}^M$), and terminates.

The rationale for this decoding algorithm is as follows. Step 1 uses the inner code to attempt to correct the errors in each symbol of the outer code. If each resulting symbol is indeed error-free, then, since the outer code is a systematic code, re-encoding the first $R_O m$ symbols (Step 2) should result in the observed decoder output (Step 3). On the other hand, if the inner codes do not succeed in correcting all symbols for the outer code, but there are fewer than $(1 - R_O)m/2 = N \log^{-9/2} N$ errors in these $m = N \log^{-3} N$ symbols, then the RS outer decoder succeeds in correcting all the outer code symbols (Step 4). Hence an error occurs only if there are $N \log^{-9/2} N$ or more errors in the outer code. The probability of this event can be bounded from above, as shown in the proof of Theorem 1 below.

Proof of Theorem 1 By the polar code exponent of [5] and our specific choice of $\beta = 3/8$, for large enough n the probability that any specific inner code fails is at most $\exp(-n^\beta) = \exp(-\log^{9/8} N)$. As noted above, our code construction fails only if $N \log^{-9/2} N$ or more of the $m = N \log^{-3} N$ inner codes fail. Hence the probability of error is bounded as

$$
\begin{aligned}
P_e &\le \binom{N \log^{-3} N}{N \log^{-9/2} N} \left(\exp\left(-\log^{9/8} N\right)\right)^{N \log^{-9/2} N} \\
&\le \exp\left(\frac{N}{\log^3 N} H_b\left(\frac{1}{\log^{3/2} N}\right)\right) \exp\left(-\frac{N}{\log^{27/8} N}\right)
\end{aligned}
$$

where the second inequality is due to Stirling's approximation. Next we note that $\lim_{\epsilon\to 0} H_b(\epsilon)/\epsilon^\alpha = 0$ for every $\alpha \in [0, 1)$. In particular, by choosing $\alpha = 1/2$, we obtain,

$$P_e \le \exp\left(N\left(\log^{-3} N\right)\left(\log^{-3/4} N\right) - \left(N \log^{-27/8} N\right)\right)$$
$$< \exp\left(\Theta(N \log^{-27/8} N)\right)$$

for large enough N. Finally, we see that this construction is capacity achieving since the inner codes and outer code are constructed at rates approaching channel capacity and 1 respectively, as N grows without bound.

Notice here that with the above choice of parameters n and m, the expected number of errors in the received codeword for the outer codeword approaches zero. Therefore, with a high probability, we receive the transmitted codeword without any error. We exploit this fact in showing that the average complexity of the decoding algorithm is dominated by the complexity of the verification step. Our inner decoder decodes each of the $\Theta(N/\log^3 N)$ inner codes using the standard polar code decoder. By the standard polar code *successive cancellation* decoding procedure [1], the computational complexity of decoding each inner code is $\mathcal{O}(\log^3 N \log(\log^3 N))$, which equals $\mathcal{O}(\log^3 N \log\log N))$. Since there are $\Theta(N/\log^3 N)$ inner codes, the overall decoding complexity of the inner code decoders is $\mathcal{O}(N \log\log N))$. This is dominated by the next decoding step, and hence we neglect this.

For our code construction, the average decoding complexity of a systematic RS code can be reduced almost to its encoding complexity ($\mathcal{O}(m \log m \log q)$ binary operations), as follows.

Recall our outer decoder does the following. It first encodes the first $R_o m$ symbols and compares the output of this encoding process with the observed m symbols. If these two sequences are the same the decoder outputs the first $R_o m$ symbols as the decoded output and terminates. If not the decoder then applies standard RS decoding [7] to the observed m symbols and outputs $R_o m$ symbols.

Let P_1 denote the probability that at least one sub-block has been decoded erroneously by the polar decoder. Since $P_1 < m\exp(-n^{3/8})$ for our choice of $\beta = 3/8$, P_1 decays as

$$P_1 < \exp(\mathcal{O}((\log N)^{9/8}))(\log^{-3} N) = o(1/m). \quad (1)$$

We now consider the complexity of this decoder for the three scenarios described above.

1. *At least* $m(1 - R_o)/2$ *inner codes fail:* This happens with probability at most $\exp(-\Omega(-N \log^{27/8} N))$. This adds $\mathcal{O}(m^2 \log m \log q)$ to the decoding complexity.
2. *None of the inner codes fail:* This happens with probability at least $1-o(1/m)$. The decoder then stops after the first step, with overall decoding complexity $\mathcal{O}(m \log m \log q)$.
3. *At least one, but fewer than* $m(1 - R_o)/2$ *inner codes fail*: This happens with probability at most $o(1/m)$. The decoder stops after the second step. This adds $\mathcal{O}(m^2 \log m \log q)$ to the decoding complexity.

Thus the expected decoding complexity is $\mathcal{O}(m \log m \log q)$. Recalling our choice of parameters $m = \Theta(N/\log^3 N)$ and $q = \exp(\Theta(\log^3 N))$, this gives an expected complexity of $\mathcal{O}(N \log N)$. ■

Calculating the multiplicative factors hidden by the Landau notation would require a more precise analysis of polar code complexity than those available to date.To get a sense of the numbers involved, we present the following examples. Each gives a rough estimate of the rate achieved under different parameter choices by abusing Landau notation and assuming that the constant multiplicative factor equals 1 everywhere.

Example 1. *For a block-length* $N = 2^{14} \approx 16,000$*, the block-length of the inner polar code is* 2744 *and that of the outer RS code is* 5*. Looking at Figure 1, the parameter of most concern then is the rate of the outer code – if the RS code has even one redundant symbol, then the difference between the overall rate and capacity is then greater than* $1/5$*, which is significant.*[6]

Example 2. *For a block-length* $N = 2^{20} \approx 10^6$*, the block-length of the inner polar code is about* $8,000$ *and that of the outer RS code is* 125*. In this case the rate of our outer code should equal* $1 - 2/\sqrt{8000} \approx 44/45$*. This suggests that about* 3 *of* 125 *symbols of the RS code should be redundant, which is perhaps more tolerable. While a block-length of* 10^6 *might seem excessive, we note that it is not inconceivable from a computational perspective when the resulting delay is acceptable. Both encoding and (with high probability) decoding can be implemented via finite field FFTs. The Cooley-Tukey algorithm [16] for FFTs allows for a parallelized implementation on multiple processors. Given current trends in computer hardware, it is not inconceivable that* 100 *different cores can each handle FFTs of length* $10,000$*, and combine them appropriately to get the overall transforms.*

Discussion of Corollary 1: Observe that outer code only operates on the message bits and can be chosen independently of the channel under consideration. Therefore, at each transmitter in a multiple-input multiple-output channel, the systematic RS code can be applied to each message independently as an outer code. Next, an inner code can be chosen to match the given channel and can be applied to the codewords from the outer code as earlier. Correspondingly, at each decoder, the overall code can be decoded by first decoding the inner code, and then using the outer code to correct failures in the inner code. Finally, noting that polar codes have been shown to be optimal for certain multiuser channels [2], and the probability of error decays in a manner similar to the single-user channel, the proof of Corollary 1 follows.

V. Code Design for Network Source Coding

Using a concatenation based construction similar to the previous section, we next show that the error probability for network source coding may be similarly reduced. As earlier, we outline the strategy for a simple network first.

A. Source Coding with Side Information at the decoder

Consider a point-to-point source coding system with side information at the decoder. The source sequence $\mathbf{u}^M \in \mathbb{F}^M$ is observed at the encoder and is demanded losslessly at the

[6] We thank the anonymous reviewer for pointing out these numbers.

decoder. In addition, the decoder also observes side information $\mathbf{y}^M \in \mathbb{F}^M$. The vector $((u_1, y_1), (u_2, y_2), \ldots, (u_M, y_M))$ is drawn i.i.d. from a joint probability mass function $p_{UY}(\mathbf{u}^M, \mathbf{y}^M) = \Pi_{i=1}^M p_{UY}(u_i, y_i)$. For this system, it is known that polar codes can asymptotically achieve the optimal rate $H(U|Y)$ with complexity $\mathcal{O}(M \log M)$ and probability of error $\exp(-\Omega(2^{M^\beta}))$ for every $\beta < 1/2$ [3].

Borrowing the parameters from the previous construction, fix $n = \log^3 M$ and $m = M \log^{-3} M$ to be the input block-lengths for the inner and outer codes. The outer code is chosen to be a systematic RS code. The concatenated code construction for this case is similar to that for channel coding, except for a few differences.

1) Encoder: Let the input blocklength for the encoder be $M = nm$. Let $f_I : \mathbb{F}_2^n :\to \mathbb{F}_2^{nR_I}$ be a polar code for this system that operates at a rate R_I and on a blocklength n and let $g_I : \mathbb{F}_2^{nR_I} \times \mathbb{F}_2^n :\to \mathbb{F}_2^n$ be the corresponding decoder. Let $R_O = (1 - 2/n^{4\beta/3})$. Let the outer code be a systematic R-S code defined via the mapping $f_o : \mathbb{F}_{2^n}^m :\to \mathbb{F}_{2^n}^{m/R_O}$. Let g_o be the corresponding decoder. Note that since f_o is chosen to be a systematic code, $f_o(\mathbf{u}^M) = (\mathbf{u}^M, \widetilde{f}_o(\mathbf{u}^M))$ for some function $\widetilde{f}_o$. The concatenated code is now defined through the mapping $f : \mathbb{F}_2^M :\to \mathbb{F}_2^{MR_I} \times \mathbb{F}_2^{M(1-R_O)}$ where, for each $\mathbf{u} \in \mathbb{F}_2^M$, $f(\mathbf{u}^M) \triangleq (f_I(u_1^n), f_I(u_{n+1}^{2n}), \ldots f_I(u_{M-n+1}^M), \widetilde{f}_o(\mathbf{u}^M))$.

2) Decoder: The decoder $g : (\mathbb{F}_2^{MR_I} \times \mathbb{F}_2^{M(1-R_O)}) \times \mathbb{F}_2^M :\to \mathbb{F}_2^M$ first decodes the polar codes and then use the redundant symbols from the R-S code to correct for block errors in the polar codes. The analysis used for probability of error and encoding and decoding complexity in Section IV can be repeated to derive similar expressions even in this case.

B. General network source coding problems

Following the observation made in Corollary 1, the strategy outlined above can be extended readily to general network source coding problems. Consider a network with multiple sources $(\mathbf{u}^M(s) : s \in S)$ such that the source $\mathbf{u}^M(s)$ is demanded losslessly at all sink nodes in the set T_s. We make the assumption that there is a directed path consisting of non-zero capacity links from a source node s to each sink node in T_s. For this setup, the concatenated code consists of a systematic R-S code as the outer code and a given network source code as the inner code. The outer code is applied to each source separately to obtain a few redundant symbols at each source node in addition to the observed source sequences. The network source code is now applied to only the observed source symbols, while the redundant symbols from the outer code are transmitted from each source s to the sinks in T_s without any coding. It can be shown that if R_O is the rate for each of the outer codes, then the extra rate required on any link is at most $|S|R_O$. Finally, observing that for specific networks such as Slepian-Wolf network and Ahlswede-Körner network etc, polar codes are optimal and the error probability vanishes as $\exp(-n^\beta)$ for $\beta < 1/2$ [2], by choosing the length of the inner code and the outer code, and the rate of the outer code, in the same way as the concatenated channel code construction, Corollary 2 follows.

VI. CONCLUSION

In this work we examine the tradeoff between the computational complexity and the probability of error of polar codes. We demonstrate that using the well-studied technique of concatenation, the probability of error can be boosted to essentially optimal performance. The question of the corresponding speed of convergence of the code rates to the optimal rate-region is still an interesting open question, as it is for the original formulation of polar codes.

VII. ACKNOWLEDGEMENTS

The authors gratefully thank Emre Telatar, Satish Korada, and Rüdiger Urbanke for introducing us to the elegant work on polar codes and patiently explaining vexing details.

REFERENCES

[1] 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, Jul 2009.
[2] S. B. Korada, “Polar codes for channel and source coding,” *Thèse EPFL*, no. 4461, 2009.
[3] N. Hussami, S. B. Korada, and R. Urbanke, “Performance of polar codes for channel and source coding,” *Proceedings of the IEEE International Symposium on Information Theory, 2009*, Jul 2009.
[4] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest: Akadémi Kiadó, 1982.
[5] E. Arikan and E. Telatar, “On the rate of channel polarization,” *Proceedings of the IEEE International Symposium on Information Theory, 2009*, pp. 1493 – 1495, Jan 2009.
[6] S. Korada, E. Sasoglu, and R. Urbanke, “Polar codes: Characterization of exponent, bounds, and constructions,” *Proceedings of the IEEE International Symposium on Information Theory, 2009*, pp. 1483 – 1487, Jan 2009.
[7] S. B. Wicker, “Error control systems for digital communication and storage,” *Prentice Hall*, 1995.
[8] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[9] G. D. Forney, “Low-density parity-check codes,” Cambridge, Massachusetts, The M.I.T. Press, 1966.
[10] D. A. Spielman, “The complexity of error-correcting codes,” *Lecture Notes in Computer Science, Fundamentals of Computation Theory*, vol. 1279/1997, pp. 67–84.
[11] I. S. Reed and G. Solomon, “Polynomial codes over certain finite fields,” *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 300 – 304, Jun 1960.
[12] F. P. Preparata and D. V. Sarwate, “Computational complexity of fourier transforms over finite fields,” *Mathematics of Computation*, vol. 31, no. 139, pp. 740 – 751, Jul 1977.
[13] G. D. Forney, “Concatenated codes,” Cambridge, Massachusetts, The M.I.T. Press, 1966.
[14] P. Delsarte and P. Piret, “Algebraic constructions of shannon codes for regular channels,” *IEEE Transactions on Information Theory*, vol. 28, no. 4, pp. 593 – 599, jul 1982.
[15] T. Uyematsu, “An algebraic construction of codes for slepian-wolf source networks,” *,IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 3082 –3088, nov 2001.
[16] J. W. Cooley and J. W. Tukey, “An algorithm for the machine calculation of complex fourier series,” *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, 1965.