2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9517921

A Practical Coding Scheme for the BSC with Feedback

Ke Wu* and Aaron B. Wagner†

*Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213 USA. kew2@andrew.cmu.edu.

†School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14850 USA. wagner@cornell.edu.

***Abstract*—We provide a practical implementation of the rubber method of Ahlswede *et al.* for binary channels. The idea is to create the "skeleton" sequence therein via an arithmetic decoder designed for a particular k-th order Markov chain. For the stochastic binary symmetric channel, we show that the scheme is nearly optimal in a strong sense for certain parameters. A byproduct of the analysis is a strict enlargement of the rates for which the sphere-packing bound is known to be achievable with feedback for this channel.**

I. Introduction

We consider the binary symmteric channel with ideal feedback, both in its stochastic- and adversarial-noise forms. In the former, each bit is flipped independently with some probability p. In the latter, an omniscient adversary can flip up to a fraction f of the bits in order to disrupt the communication.

The information-theoretic limits for both forms of the channel, assuming perfect feedback, are well-known. In the adversarial case, the capacity as a function of f was determined by Zigangirov [1], building on earlier results of Berlekamp [2], [3]. For the stochastic version, the capacity equals that of the non-feedback version (e.g., [4], [5]) and likewise the high-rate error exponent, normal approximation, and moderate deviations performance are all unimproved by feedback. In fact, the third-order coding rate is unimproved by feedback [6], as is the order of the optimal "pre-factor" in front of the error exponent at high rates. Thus, at least for the stochastic version of the channel, feedback offers very little improvement in coding performance.

In general, feedback is known to simplify the coding problem even if it does not provide for improved performance. The erasure (e.g., [7, Section 17.1]), Gaussian channels [8], [9] provide striking examples of this phenomenon. For the BSC, see [10], [11] for classical and [12], [13] for recent work on devising implementable schemes using feedback.

For the adversarial symmetric channel with feedback (and arbitrary, finite alphabet size), Ahlswede *et al.* [14] proposed an explicit scheme called the *rubber method*. In the binary case, for a fixed $\ell > 2$, the message is encoded as a "skeleton" string containing no substring of ℓ consecutive zeros. The encoder then transmits this string, sending ℓ consecutive zeros to indicate that an error has occurred. For each ℓ, this scheme achieves the capacity of the adversarial channel for a certain choice of f. This scheme simplifies significantly the original achievability argument of Berlekamp [2]. For ternary and larger alphabets, the scheme is even simpler. The rubber method has since been generalized [15]–[17]. See the survey [18] for applications and different versions of the rubber method. However, the rubber coding method does not specify an explicit way to map the message sequences to skeleton sequences, which makes it unimplementable. The classical schemes of Berlekamp [2] and Schalkwijk [11] are not amenable to direct implementation, either.

Our Contribution We only consider the binary case in this paper, and we make three contributions. The first is to propose the use of arithmetic coding applied to a particular Markov chain in order to efficiently encode the message sequence into the corresponding skeleton string. This results in a practically-implementable end-to-end scheme, with only a negligible rate penalty. Our coding scheme is universal, i.e., the encoder does not require knowledge of the cross-over probability p. Moreover, it works in a streaming manner: the encoder does not need the entire message to begin encoding. The second contribution is showing that, for each ℓ, there is a special rate R_ℓ^* and crossover probability p_ℓ such that the resulting scheme is optimal with respect to the second-order coding rate and moderate deviations performance for the channel with crossover probability p_ℓ and error-exponent optimal at rate R_ℓ^* for all channels with crossover probability less than p_ℓ. We also consider the third-order coding rate and the "pre-factor" of the error exponent of the scheme. These turn out to be nearly, but not exactly optimal. See Section V for details.Our scheme has stronger optimality guarantees than Horstein [10] and Li and El Gamal [19], although these schemes are more general. The third contribution is that we strictly enlarge the set of rates for which the sphere-packing bound is known to be achievable for the BSC with feedback.

Technical Overview. We implement the rubber coding method by using arithmetic decoding to construct skeleton sequences from the message bits. This is accomplished by characterizing a Markov chain that places a uniform distribution over the set of skeleton sequences. We then characterize the performance of this scheme in the error exponent, moderate deviations, and normal approximation regimes.

In Section II we introduce our notation and provide various preliminaries. In Section III we characterize the relevant Markov chain. In Section IV we describe our coding scheme. In Section V we present our main results. The proofs are omitted due to space constraints, but are available in the full version [20].

978-1-5386-8209-8/21/$31.00 ©2021 IEEE

II. Notation and Preliminaries

Capital letters such as X or Y denote random variables. We use x^n to denote the first n bits of the sequence $x_1, \ldots, x_n$, and we use $z \| z'$ to denote the concatenation of two strings z and z'. In addition, $\lfloor x^N \rfloor_L$ denotes the truncation of x^N to the first L bits.

We use $\mathsf{Bin}(n,p)$ to denote the binomial distribution with size n and success probability p and $\mathcal{N}(\mu, \sigma^2)$ to denote the normal distribution with mean μ and variance σ^2. Moreover, $B(p)$ denotes the Bernoulli distribution with success probability p. We use $D(P\|Q)$ to denote the Kullback-Leibler divergence between distribution P and Q.

A. The Channel Model

Let $\mathsf{BSC}(p)$ denote a binary symmetric channel with crossover probability $p \in (0, \frac{1}{2})$ without feedback. That is, $\mathsf{BSC}(p)$ has input alphabet $\mathcal{X} = \{0,1\}$ and output alphabet $\mathcal{Y} = \{0,1\}$, and probability transition matrix

$$p(y|x) = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}.$$

Suppose that an encoder wishes to send a message m in a message space $\mathcal{M}$ through $\mathsf{BSC}(p)$. It first encodes the message m using an encoding function f, and sends $x^N = f(m)$ through the channel. The decoder, upon receiving y^N from the channel, runs a decoding function g on y^N to obtain m'. The pair (f,g) is called a *code* $\mathcal{C}_{N,R}$ with block length N and rate $R = \frac{\log|\mathcal{M}|}{N}$. The (average) error probability of a code $\mathcal{C}_{N,R}$ is defined as $P_e(\mathcal{C}_{N,R}) := \frac{1}{|\mathcal{M}|}\sum_{m \in \mathcal{M}} \Pr[m' \neq m]$.

The *capacity* of $\mathsf{BSC}(p)$ is well-known to be

$$C(\mathsf{BSC}(p)) = 1 - h(p),$$

where $h(\cdot) = -p\log p - (1-p)\log(1-p)$ is the binary entropy function, and the $\log$ is base-2 throughout.

We will also consider the adversarial binary symmetric channel $\mathsf{BSC}_{adv}(f)$ in which at most f fraction of transmitted bits can be adversarially flipped.

Feedback allows the encoder to see exactly what the decoder receives after each transmission and update its next transmission accordingly. In the BSC with feedback, which we denote as $\mathsf{BSC}^{fb}(p)$, the encoding function f consists of a sequence of maps $\{f_i\}_{i=1}^N$. Each f_i takes as input $m, y_1, \ldots, y_{i-1}$, and outputs x_i, the next bit to send. The decoder then runs $g(y^N)$ to obtain m'.

It is well-known that feedback does not improve the channel capacity:

$$C(\mathsf{BSC}^{fb}(p)) = 1 - h(p).$$

For the adversarial feedback BSC channel $\mathsf{BSC}^{fb}_{adv}(f)$, an upper bound on the capacity was first shown by Berlekamp [2]. He also gives a lower bound that coincides with the upper bound when $f \geq \frac{3-\sqrt{5}}{4}$. A lower bound that coincides with the upper bound for $f < \frac{3-\sqrt{5}}{4}$ was given by Zigangirov [1], thus determining the capacity for $\mathsf{BSC}^{fb}_{adv}(f)$:

$$C(\mathsf{BSC}^{fb}_{adv}(f)) = \begin{cases} 1 - h(f) & \text{if } 0 \leq f \leq \frac{3-\sqrt{5}}{4}, \\ (1-3f)\log\frac{1+\sqrt{5}}{2} & \text{if } \frac{3-\sqrt{5}}{4} < f \leq 1. \end{cases}$$

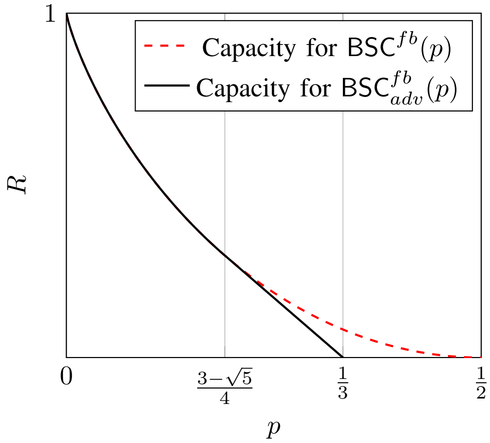

Fig. 1: Capacity for BSC with feedback and adversarial BSC with feedback.

We say that a code $\mathcal{C}$ for the $\mathsf{BSC}^{fb}_{adv}(f)$ is *admissible* if $\mathcal{C}$ can correct any error pattern with error fraction at most f. We say that a sequence of codes $\{\mathcal{C}_{N,R}\}_N$ for $\mathsf{BSC}^{fb}(p)$ is *admissible* if the error probability $P_e(\mathcal{C}_{N,R})$ tends to 0 as N goes to infinity.

B. Markov Chains

Definition 1. A discrete stochastic process $\{X_i\}$ is said to be an $(\ell-1)$-th order Markov chain if for any i,

$$\begin{aligned} &\Pr[X_i = x_i | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}] \\ =&\Pr[X_i = x_i | X_{i-\ell+1} = x_{i-\ell+1}, \ldots, X_{i-1} = x_{i-1}], \end{aligned}$$

for all $x_1, \ldots, x_i \in \mathcal{X}$.

C. Rubber Method

Here we briefly present the rubber method for $\mathsf{BSC}^{fb}_{adv}(f)$ [14]. Let $\mathcal{A}_\ell^{N'}$ denote the set of binary sequences of length N' with no ℓ consecutive zeros. Such sequences are called *skeleton sequences*. The sender chooses a skeleton sequence $x^{N'} \in \mathcal{A}_\ell^{N'}$ and the decoder's goal is to recover that sequence correctly. The idea is that the encoder can use ℓ consecutive zeros to signal an error. Specifically, we have

- Decoding g_R: the decoder maintains a stack of received bits, which begins empty. Whenever the decoder receives a bit, it inserts the received bit onto the stack and checks if there are consecutive ℓ zeros in the stack. If yes, it removes these ℓ zeros as well as the bit before these consecutive ℓ zeros from the stack. Finally, it truncates the output to N' bits.
- Encoding f_R: if the decoder's current stack is a prefix of $x^{N'}$, then send the next bit in $x^{N'}$. Otherwise send a 0. If $x^{N'}$ has been sent in its entirety, then send 1 for all remaining time steps.

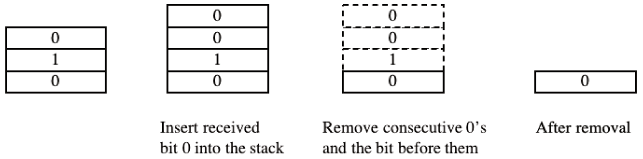

Fig. 2: The decoder's stack with $\ell = 2$.

Proposition 2. *For skeleton sequence set $\mathcal{A}_\ell^{N'}$ and block length N, a code constructed using the rubber method is admissible for $\mathsf{BSC}_{adv}^{fb}(f)$ if*

$$N' + (\ell + 1)fN \leq N. \tag{1}$$

Proof. See Section 2.2 of [21]. □

Example 3. Suppose the encoder chooses $x = 011010 \in \mathcal{A}_2^6$ and the maximum fraction of adversarial errors is $f = 1/3$.

Suppose the first three bits the decoder receives are 010, which is not a prefix of x. The encoder then sends 0 and suppose decoder sees 0100. The decoder then erases the last three bits (the consecutive zeros and the one before them) and its stack becomes 0. This is now a prefix of x and the encoder would thus resend the second bit in x, which is 1. See Figure 2.

D. Shannon–Fano–Elias Code and Arithmetic Coding

The Shannon–Fano–Elias code compresses a source sequence with known distribution to near-optimal length. It uses the cumulative distribution function $F(x)$ to allot codewords. For a random variable $X \in \{1, 2, \ldots, M\}$ with distribution p, the codeword is $\lfloor \bar{F}(x) \rfloor_{l(x)}$ where

$$\bar{F}(x) = \sum_{a<x} p(a) + \frac{1}{2}p(x),$$

lies between $F(x)$ and $F(x+1)$ and $l(x) = \lceil \log \frac{1}{p(x)} \rceil + 1$. In addition, the Shannon-Fano-Elias code is prefix-free. That is, no codeword is a prefix of any other.

Arithmetic coding is an algorithm for efficiently computing the Shannon–Fano–Elias codeword for sequences given a method for computing the probability of the next symbol given the past (e.g., [22, Ch. 4]).

E. Constant Recursive Sequences and the Perron–Frobenius Theorem

Lemma 4 (Theorem 2.3.6, [23])**.** *A sequence $A(n)$ is an order-d* constant-recursive sequence *if for all $n \geq d + 1$,*

$$A(n) = c_1 A(n-1) + c_2 A(n-2) + \cdots + c_d A(n-d).$$

The n-th term $A(n)$ in the sequence must be of the form

$$A(n) = k_1(n)\lambda_1^n + k_2(n)\lambda_2^n + \cdots + k_{d'}(n)\lambda_{d'}^n,$$

where λ_i is a root with multiplicity d_i of the polynomial

$$\lambda^d - c_1\lambda^{d-1} - \cdots - c_d,$$

and $k_i(n)$ is a polynomial with degree $d_i - 1$.

Definition 5 ((8.3.16), [24])**.** A matrix M is a *positive (non-negative)* matrix if every entry of M is positive (non-negative).

A non-negative square matrix M is *primitive* if its k-th power is positive for some natural number k.

Lemma 6 (Perron–Frobenius Theorem, Page 674, [24])**.** *If M is a primitive matrix, then M has a positive real eigenvalue λ^* such that all other eigenvalues λ_i have absolute value $|\lambda_i| < |\lambda^*|$. Moreover, λ^* is a simple eigenvalue and its corresponding column and row eigenvectors are positive .*

See [24, Ch. 8] for further detail about the Perron-Frobenius Theorem.

III. A Key Markov Chain

In this section we show that the stochastic process that is uniformly distributed over $\mathcal{A}_\ell^N$ is a Markov Chain and that we can efficiently compute the distribution of this Markov Chain.

Recall that $\mathcal{A}_\ell^N$ denotes the set of binary sequences of length N with no consecutive ℓ zeros. Let $A_\ell(N) = |\mathcal{A}_\ell^N|$ and let $A_\ell(z)$ denote the number of allowable sequences in $\mathcal{A}_\ell^N$ that begin with z for any binary sequence z.

Lemma 7. *Let λ_ℓ^* be the unique real solution that lies in $(1, 2)$ of*

$$\lambda^\ell = \lambda^{\ell-1} + \lambda^{\ell-2} + \cdots + 1. \tag{2}$$

Then $\lim_{N\to\infty} \frac{|\mathcal{A}_\ell^N|}{\lambda_\ell^{*N}}$ exists and is positive and finite.

Note that Lemma 7 implies that

$$\lim_{N\to\infty} \frac{1}{N} \log |\mathcal{A}_\ell^N| = \log \lambda_\ell^*.$$

Lemma 8. *The stochastic process that is uniformly distributed over $\mathcal{A}_\ell^N$ is an $(\ell - 1)$-th order Markov Chain.*

The resulting Markov Chain is:

- For any $x \in \{0, 1\}^{\ell-1}$,

$$\Pr[X_1, \ldots, X_{\ell-1} = x] = \frac{A_\ell(N - \ell + 1 + \alpha)}{A_\ell(N)},$$

  if x ends with α consecutive zeros.
- For $i \geq \ell$, for any $z \in \{0, 1\}^{\ell-1}$,

$$\begin{aligned}&\Pr[X_i = 1 | X_{i-\ell+1}, \ldots, X_{i-1} = z]\\ =&\frac{A_\ell(N-i)}{A_\ell(N-i+\alpha+1) - \sum_{k=0}^{\alpha-1} A_\ell(N-i+k+1)},\end{aligned}$$

  if z ends with α consecutive zeros.

To compute the probability of the next symbol in the string given the past, we only need to compute $A_\ell(N)$ for various values of N. This can be computed using $A_\ell(N) = k_1\lambda_1^N + k_2(N)\lambda_2^N + \cdots + k_{\ell'}(N)\lambda_{\ell'}^N$ where λ_i are the roots of equation (2) and $c_1, k_1(N), \ldots, k_{\ell'}(N)$ can be determined by the initial conditions $A_\ell(1) = 2, \ldots, A_\ell(\ell - 1) = 2^{\ell-1}, A_\ell(\ell) = 2^\ell - 1$.

Note that when N is large, $A_\ell(N)$ is well-approximated as $A_\ell(N) \approx k_1\lambda_\ell^{*N}$. Under this approximation the Markov Chain becomes time-invariant.

Example 9. Consider the case $\ell = 2$. That is, we forbid two consecutive zeros in the skeleton sequence. Then the characteristic polynomial is $\lambda^2 - \lambda - 1 = 0$. The two roots are $\lambda_1 = \frac{1+\sqrt{5}}{2}$ and $\lambda_2 = \frac{1-\sqrt{5}}{2}$ respectively. The initial condition is $A_\ell(1) = 2, A_\ell(2) = 3$. Therefore $A_\ell(N) = k_1\lambda_1^N + k_2\lambda_2^N$ where $k_1 = \frac{3+\sqrt{5}}{2\sqrt{5}}$, $k_2 = \frac{\sqrt{5}-3}{2\sqrt{5}}$. See also [4, Ex. 4.7]

IV. A Practical Coding Scheme

In this section we combine arithmetic coding and the rubber method to give an efficient feedback code for $\mathsf{BSC}_{adv}^{fb}(f)$ and $\mathsf{BSC}^{fb}(p)$. First we describe a modified version of arithmetic coding that will be used in our scheme. Consider the following pair of algorithms $(\mathsf{Decom}_\ell, \mathsf{Com}_\ell)$:

Algorithm 10. $(\mathsf{Decom}_\ell, \mathsf{Com}_\ell)$

Let $L = \lceil \log |\mathcal{A}_\ell^N| \rceil$. Let $\{X_i\}_{i=1}^N$ be a stochastic process that is uniformly distributed over $\mathcal{A}_\ell^N$. Let (A_C, A_D) where $A_C : \mathcal{A}_\ell^N \mapsto \{0,1\}^{L+1}$ and $A_D : \{0,1\}^{L+1} \mapsto \mathcal{A}_\ell^N \cup \{\perp\}$ be the compression and decompression algorithms for arithmetic coding applied to $\{X_i\}_{i=1}^N$, where the decompressor outputs $\perp$ if its input is not a valid codeword. Let L' be any integer such that $L' \leq L - 3$.

$\mathsf{Decom}_\ell(m) : \{0,1\}^{L'} \mapsto \mathcal{A}_\ell^N$

1) Run the decompress algorithm $A_D(m\|m')$ for all possible $m' \in \{0,1\}^{L+1-L'}$. Let the first non-$\perp$ output be $A_D(m\|m') = x^N$. If there's no such x^N, set x^N to be a random sequence in $\mathcal{A}_\ell^N$.
2) Output x^N.

$\mathsf{Com}_\ell(x^N) : \mathcal{A}_\ell^N \mapsto \{0,1\}^{L'}$:

1) Output $\lfloor A_C(x^N) \rfloor_{L'}$.

Lemma 11. *The pair of algorithms* $(\mathsf{Decom}_\ell, \mathsf{Com}_\ell)$ *described in Algorithm 10 satisfies*

$$\mathsf{Com}_\ell(\mathsf{Decom}_\ell(m)) = m, \forall m \in \{0,1\}^{L'}.$$

Now we describe the construction of our overall scheme:

Construction 12. The encoding and decoding of $\mathcal{C}_{\ell,N,R}$ are as follows:

Encoding:

- Let m^{NR} be a message source of length NR. Find the minimum natural number N' such that $\lceil \log |\mathcal{A}_\ell^{N'}| \rceil \geq NR + 3$.
- Run $\mathsf{Decom}_\ell(m)$ and denote the output as $x^{N'}$. Let $x^{N'}$ be the skeleton sequence and send it through the feedback channel using the rubber method.

Decoding:

- Let y^N be the sequence received from the feedback channel. Run the decoding algorithm of the rubber method on y to get $\widetilde{x}^{N'}$. If $\widetilde{x}^{N'} \notin \mathcal{A}_\ell^N$, set $\widetilde{x}^{N'}$ to be a random skeleton sequences in $\mathcal{A}_\ell^{N'}$.
- Otherwise, output $m' = \mathsf{Com}_\ell(\widetilde{x}^{N'})$.

Proposition 13. *The code* $\mathcal{C}_{\ell,N,R}$ *in Construction 12 is admissible for the* $\mathsf{BSC}_{adv}^{fb}(f)$ *if* $N' \leq (1-(\ell+1)f)N$.

ℓ	$\log \lambda_\ell^*$	p_ℓ	R_ℓ^*
2	0.6942	0.1910	0.2965
3	0.8791	0.0804	0.5965
4	0.9468	0.0362	0.7754

TABLE I: Numerical results of $\log \lambda_\ell^*$, tangent points p_ℓ and tangent rates R_ℓ^*

Note that in the first step of encoding, we can find N' simply by computing $\mathcal{A}_\ell^{\widetilde{N}}$ for $\widetilde{N} = NR+3, \ldots, 2NR+6$ since $2^{\frac{\widetilde{N}}{2}} \leq |\mathcal{A}_\ell^{\widetilde{N}}| \leq 2^{\widetilde{N}}$.

We further note that the above coding scheme also works for stochastic feedback BSC channel $\mathsf{BSC}^{fb}(p)$:

Proposition 14. *The sequence of codes* $\{\mathcal{C}_{\ell,N,R}\}_N$*, each of which is constructed as in Construction 12, is admissible for the* $\mathsf{BSC}^{fb}(p)$ *if* $R < R_\ell(p) = (1-(\ell+1)p)\log\lambda_\ell^*$.

V. Main Results

We now show that, for certain parameters, our codes achieve the capacity and the optimal error-exponent, second-order rate, and moderate deviations constant for certain parameters.

A. Capacity

Theorem 15. *For any integer* $\ell \geq 2$, $R_\ell(p)$ *is tangent to* $C(\mathsf{BSC}^{fb}(p))$. *For* $p_\ell = \frac{1}{1+2^{(\ell+1)\log\lambda_\ell^*}}$,

$$R_\ell(p_\ell) = C(\mathsf{BSC}^{fb}(p_\ell)).$$

That is, for any $\epsilon > 0$, *the sequence of codes* $\{\mathcal{C}_{\ell,N,R}\}_N$ *as constructed in Construction 12 is admissible for* $\mathsf{BSC}^{fb}(p_\ell)$ *with* $R = C(\mathsf{BSC}^{fb}(p_\ell)) - \epsilon$.

We call p_ℓ the *tangent points* and $R_\ell^* = R_\ell(p_\ell)$ the *tangent rates*. The tangent points p_ℓ, tangent rates R_ℓ^*, and $\log\lambda_\ell^*$ values for different ℓ are listed in Table I.

The function $R_\ell(p)$ for different ℓ is plotted in Figure 3. That the rubber method would achieve the capacity of the $\mathsf{BSC}^{fb}(p_\ell)$ is implicit in [21]. We consider three more-refined performance measures.

B. Error-exponent

Lemma 16 (Sphere-packing bound with pre-factor [6], [25])**.** *Let* $\{\mathcal{C}_{N,R}\}_N$ *be a sequence of codes for the* $\mathsf{BSC}^{fb}(p)$*, each with rate* $R < C(\mathsf{BSC}^{fb}(p))$. *Let* $q \in (0, \frac{1}{2})$ *s.t.* $R = 1 - h(q)$. *Let* $E_{sp}(R) = D(B(q)\|B(p))$ *and* $E'_{sp}(R)$ *be the slope of this function at* R. *Then the error probability* $P_e(\mathcal{C}_{N,R})$ *satisfies*

$$P_e(\mathcal{C}_{N,R}) \geq \frac{K_1}{N^{\frac{1}{2}(1+|E'_{sp}(R)|)}} e^{-NE_{sp}(R)},$$

where K_1 *is a positive constant depending on* R.

Theorem 17. *For any fixed* $\ell \geq 2$, *consider the sequence of codes* $\{\mathcal{C}_{\ell,N,R_\ell^*}\}_N$ *at the tangent rate* R_ℓ^*. *That is,* $R_\ell^* = R_\ell(p_\ell) = 1 - h(p_\ell)$. *Then for the* $\mathsf{BSC}^{fb}(p)$ *with* $p < p_\ell$, $\{\mathcal{C}_{\ell,N,R_\ell^*}\}_N$ *at rate* R_ℓ^* *achieves optimal error exponent*

$$P_e(\mathcal{C}_{\ell,N,R_\ell^*}) \leq O\left(\frac{1}{\sqrt{N}}\right) e^{-N\cdot E_{sp}(R)}.$$

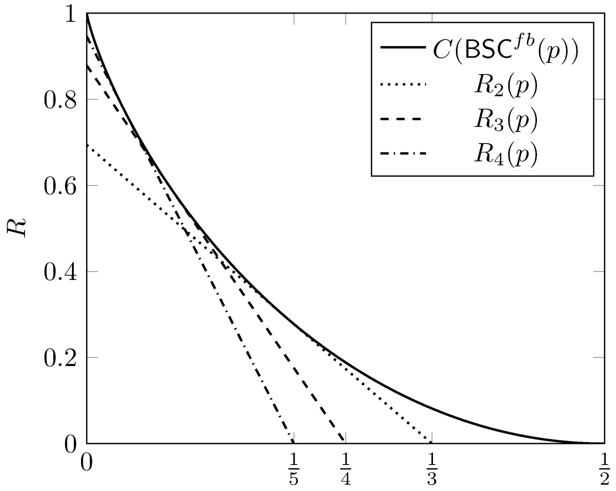

Fig. 3: $R_\ell(p)$ for different ℓ.

In particular,

$$\lim_{N\to\infty} -\frac{1}{N}\log P_e(\mathcal{C}_{\ell,N,R^*_\ell}) = E_{sp}(R).$$

Remark 18. The "pre-factor" order achieved by our scheme is $O(\frac{1}{\sqrt{N}})$, which is slightly worse than the optimal order of $O(\frac{1}{N^{\frac{1}{2}(1+|E'_{sp}(R)|)}})$ in Theorem 16. Interestingly, for the binary erasure channel (BEC), both with and without feedback, the optimal pre-factor is $O(\frac{1}{\sqrt{N}})$ [6, Theorem 2]. Rubber coding attempts to emulate a BEC using the BSC, which might explain this connection. A similar gap from strict optimality occurs in the second-order coding rate results to follow. Making the connection between rubber coding and the BEC more precise is an interesting topic for future study.

Theorem 17 indicates that our coding scheme $\mathcal{C}_{2,N,R^*_\ell}$ at the tangent rate achieves the sphere-packing bound for $\mathsf{BSC}^{fb}(p)$ for any $p < p_2$. Previously the sphere-packing bound was only known to be achievable with feedback for rates above *critical rate with feedback* [26]. We note that $R^*_2 = R_2(p_2)$ is below the critical rate with feedback for some cross-over probabilities p (see Fig. 4). Thus a byproduct of this work is that we strictly enlarge the set of rates for which the sphere-packing bound is known to be achievable for the BSC with feedback.

C. Second-order Rate

Lemma 19 (Second-order coding rate: Theorem 15, [27])**.** *Given a block length N and an ϵ such that $0 < \epsilon < 1$, the largest possible rate of a code for the $\mathsf{BSC}^{fb}(p)$ with error probability less than or equal to ϵ is*

$$C - \frac{1}{\sqrt{N}}\sqrt{p(1-p)\log^2\frac{1-p}{p}}\,\Phi^{-1}(1-\epsilon) + \frac{\log N}{2N} + o(1),$$

where Φ denotes the standard Gaussian distribution.

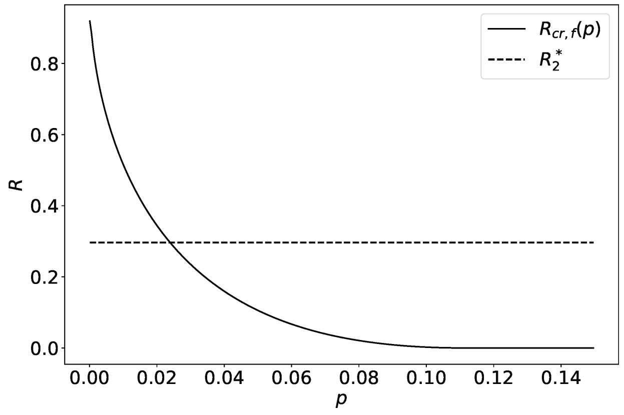

Fig. 4: The critical rate with feedback $R_{cr,f}$ compared to the tangent rate $R^*_2 = R_2(p_2)$, as a function of the crossover probability p.

Theorem 20. *For any fixed $\ell \geq 2$, consider the $\mathsf{BSC}^{fb}(p)$ with cross-over probability $p = p_\ell$. Fix $\epsilon \in (0,1)$, and let $R(N,\epsilon)$ denote the largest possible rate R such that $\mathcal{C}_{\ell,N,R(N,\epsilon)}$ has error probability at most ϵ, and let C denote the capacity of the $\mathsf{BSC}^{fb}(p)$. Then for large N,*

$$\begin{aligned} &R(N,\epsilon) \\ &\geq C - \frac{1}{\sqrt{N}}\sqrt{p(1-p)\log^2\frac{1-p}{p}}\,\Phi^{-1}(1-\epsilon) - O\left(\frac{1}{N}\right). \end{aligned}$$

Remark 21. Note the $\log N/N$ term is "missing" from the expansion in Theorem 20. See Remark 18.

D. Moderate Deviations

Lemma 22 (Moderate deviations, Corollary 1, [28])**.** *For any sequence of real numbers ϵ_N s.t. $\epsilon_N \to 0$ as $N \to \infty$ and $\epsilon_N\sqrt{N} \to \infty$ as $N \to \infty$, for any sequence of codes $\{\mathcal{C}_{N,R_N}\}_N$ for the $\mathsf{BSC}^{fb}(p)$ such that $R_N \geq C(\mathsf{BSC}^{fb}(p)) - \epsilon_N$, we have*

$$\liminf_{N\to\infty} \frac{1}{N\epsilon_N^2}\log P_e(\mathcal{C}_{N,R_N}) \geq -\frac{1}{2p(1-p)\log^2\frac{1-p}{p}}.$$

Theorem 23. *Fix any $\ell \geq 2$. Let C be the capacity of the $\mathsf{BSC}^{fb}(p_\ell)$. For any sequence of real numbers ϵ_N s.t. $\epsilon_N \to 0$ as $N \to \infty$ and $\epsilon_N\sqrt{N} \to \infty$ as $N \to \infty$, consider the sequence of codes $\{\mathcal{C}_{\ell,N,R_N}\}_N$ such that $R_N = C - \epsilon_N$. Let $P_e(\mathcal{C}_{\ell,N,R_N})$ denote the average error probability of $\mathcal{C}_{\ell,N,R_N}$ over the $\mathsf{BSC}^{fb}(p_\ell)$. Then*

$$\lim_{N\to\infty} \frac{1}{N\epsilon_N^2}\log P_e(\mathcal{C}_{\ell,N,R_N}) = -\frac{1}{2p(1-p)\log^2\frac{1-p}{p}}.$$

Acknowledgment

This research was supported by the US National Science Foundation under grant CCF-1956192 and the US Army Research Office under grant W911NF-18-1-0426.

References

[1] K. Zigangirov, "On the number of correctable errors for transmission over a binary symmetrical channel with feedback," *Problemy Peredachi Informatsii*, vol. 12, no. 2, pp. 3–19, 1976.

[2] E. R. Berlekamp, "Block coding for the binary symmetric channel with noiseless, delayless feedback," in *Error-Correcting Codes*, H. B. Mann, Ed., the Mathematics Research Center, United States Army at the University of Wisconsin, Madison. Wiley New York, May 1968, pp. 61–68.

[3] ——, "Block coding with noiseless feedback," *PhD. Thesis*, 1964.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.

[5] C. Shannon, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, 1956.

[6] Y. Altuğ and A. B. Wagner, "On exact asymptotics of the error probability in channel coding: symmetric channels," *IEEE Trans. Inf. Theory*, vol. 67, no. 2, pp. 844–868, 2021.

[7] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[8] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. 12, no. 2, pp. 172–182, 1966.

[9] J. Schalkwijk, "A coding scheme for additive noise channels with feedback–II: Band-limited signals," *IEEE Trans. Inf. Theory*, vol. 12, no. 2, pp. 183–189, 1966.

[10] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 9, no. 3, pp. 136–143, 1963.

[11] J. Schalkwijk, "A class of simple and optimal strategies for block coding on the binary symmetric channel with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 17, no. 3, pp. 283–287, 1971.

[12] H. Yang and R. D. Wesel, "Finite-blocklength performance of sequential transmission over BSC with noiseless feedback," in *Proc. IEEE Intl. Symp. on Inf. Theory (ISIT)*, 2020, pp. 2161–2166.

[13] A. Antonini, H. Yang, and R. D. Wesel, "Low complexity algorithms for transmission of short blocks over the BSC with full feedback," in *Proc. IEEE Intl. Symp. on Inf. Theory (ISIT)*, 2020, pp. 2173–2178.

[14] R. Ahlswede, C. Deppe, and V. Lebedev, "Non-binary error correcting codes with noiseless feedback, localized errors, or both," in *Proc. IEEE Intl. Symp. on Inf. Theory (ISIT)*, 2006, pp. 2486–2487.

[15] V. S. Lebedev, "Coding with noiseless feedback," *Problems of Information Transmission*, vol. 52, no. 2, pp. 103–113, 2016.

[16] C. Deppe, V. Lebedev, G. Maringer, and N. Polyanskii, "Coding with noiseless feedback over the Z-channel," in *International Computing and Combinatorics Conference*. Springer, 2020, pp. 98–109.

[17] C. Deppe, V. Lebedev, and G. Maringer, "Bounds for the capacity error function for unidirectional channels with noiseless feedback," *Theoretical Computer Science*, 2020.

[18] C. Deppe, G. Maringer, and V. Lebedev, "How to apply the rubber method for channels with feedback," in *Algebraic and Combinatorial Coding Theory (ACCT)*, 2020, pp. 1–6.

[19] C. T. Li and A. El Gamal, "An efficient feedback coding scheme with low error probability for discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 61, no. 6, pp. 2953–2963, 2015.

[20] K. Wu and A. B. Wagner, "A practical coding scheme for the BSC with feedback." [Online]. Available: https://arxiv.org/pdf/2102.02358.pdf

[21] R. Ahlswede, C. Deppe, and V. Lebedev, "Non–binary error correcting codes with noiseless feedback, localized errors, or both," *Annals of European Academy of Sciences*, no. 1, pp. 285 – 309, 2005. [Online]. Available: https://www.math.uni-bielefeld.de/ahlswede/homepage/public/181.pdf

[22] K. Sayood, *Introduction to Data Compression*. Morgan Kaufmann, 2017.

[23] P. Cull, M. Flahive, and R. Robson, *Difference Equations: From Rabbits to Chaos*. Spring-Verlag NY Inc., 2005.

[24] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Siam, 2000, vol. 71.

[25] P. Elias, "Coding for two noisy channels," in *3rd London Symp. on Inf. Theory*, 1955, pp. 61–76.

[26] K. Zigangirov, "Upper bounds for the error probability for channels with feedback," *Problemy Peredachi Informatsii*, vol. 6, no. 2, pp. 87–92, 1970.

[27] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Feedback in the non-asymptotic regime," *IEEE Trans. on Inf. Theory*, vol. 57, no. 8, pp. 4903–4925, 2011.

[28] Y. Altuğ, H. V. Poor, and S. Verdú, "On fixed-length channel coding with feedback in the moderate deviations regime," in *Proc. IEEE Intl. Symp. on Inf. Theory (ISIT)*, 2015, pp. 1816–1820.