An Upper Bound on the Reliability Function of the DMC with or without Mismatch

Anelia Somekh-Baruch
Faculty of Engineering
Bar-Ilan University
Ramat-Gan, Israel
Email: somekha@biu.ac.il

Abstract—We derive a new upper bound on the reliability function for channel coding over discrete memoryless channels. Our bounding technique relies on two main elements: (i) adding an auxiliary genie-receiver that reveals to the original receiver a list of codewords including the transmitted one, which satisfy a certain type property, and (ii) partitioning (most of) the list into subsets of codewords that satisfy a certain pairwise-symmetry property, which facilitates lower bounding of the average error probability by the pairwise error probability within a subset. We compare the obtained bound to the Shannon-Gallager-Berlekamp straight-line bound, the sphere-packing bound, and an amended version of Blahut’s bound. Our bound is shown to be at least as tight as for all rates, with cases of stricter tightness in a certain range of low rates, compared to all three aforementioned bounds.

Our derivation is performed in a unified manner which is valid for any rate, as well as for a wide class of additive decoding metrics, whenever the corresponding zero-error capacity is zero. We also present a dual form of the bound, and discuss a looser bound of a simpler form, which is analyzed for the case of the binary symmetric channel with maximum likelihood decoding.

I. INTRODUCTION

It is well known that the average error probability in coding for the discrete memoryless channel (DMC) can be made to vanish exponentially fast with the code block-length \( n \), if the rate of the code is strictly lower than the channel capacity. The reliability function signifies exactly how fast the exponential decay can be, as a function of the code rate \( R \). The problem of finding a single-letter expression for the reliability function has been studied extensively since the early days of information theory (see, e.g., [1]–[5]).

For a certain range of high rates, the random-coding lower bound coincides with the sphere-packing upper bound [1], [2], establishing the exact value of the reliability function curve above a certain critical rate. The exact value of the reliability function for zero-rate codes (at \( R = 0^+ \)) is also known, and was characterized by Shannon, Gallager and Berlekamp [3] who obtained a tight upper bound that matches the expurgated lower bound at zero rate [1], for channels whose zero-error capacity equals zero. Another important result of [3], which holds for the same class of channels, is referred to as the straight-line upper bound on the reliability function at low rates, which is obtained by the tangential line to the sphere packing bound curve that meets the \( R = 0 \) axis at the zero-rate reliability point.

In [5] an upper bound on the reliability function was proposed, that contained a gap in the proof, which was revisited in [6]. The correction was based on Blinkovsky’s [7] idea of using a Ramsey-theoretic result by Komlós [8]. This yielded an upper bound applicable to general DMCs.

For certain special cases of channels, tighter upper bounds on the reliability function have been derived sometimes enlarging the range of rates for which it is known, see, e.g., [9]. However, deriving a single-letter formula for the reliability function for a certain range of low rates remains an open problem, and so is the question of understanding the structure of low rate good codes. For a survey of the subject of the channel coding reliability function see [10] and references therein.

Certain works have extended the study of lower and upper bounds on the reliability function of the DMC to the more general setup, where the decoding metric is not necessarily optimal, a.k.a. mismatched decoding (see [11]–[17]).

II. A FORMAL STATEMENT OF THE PROBLEM

Consider transmission over a memoryless channel described by a conditional probability rule \( W(y|x) \), with input \( x \in X \) and output \( y \in Y \) finite alphabets \( X \) and \( Y \). We define \( W^n(y|x) = \prod_{k=1}^{n} W(y_k|x_k) \) for input/output sequences \( x = (x_1,\ldots,x_n) \in X^n \) and \( y = (y_1,\ldots,y_n) \in Y^n \). The corresponding random variables are denoted by \( X \) and \( Y \).

An encoder maps a message \( m \in \{1,\ldots,M_n\} \) to a channel input sequence \( x_m \in X \), where the number of messages is denoted by \( M_n \). The message, represented by the random variable \( M \), is assumed to take values in \( \{1,\ldots,M_n\} \) equi-probably. This mapping induces an \( (n,M_n) \)-codebook \( \mathcal{C}_n = \{x_1,\ldots,x_{M_n}\} \) with rate \( R_n = \frac{1}{n} \log M_n \).

Upon observing the channel output \( y \), the decoder, who wishes to minimize the average error probability, can use the optimal ML decoding rule

\[
\hat{m} = \arg \max_{m \in \{1,\ldots,M_n\}} W^n(y|x_m) = \arg \max_{m \in \{1,\ldots,M_n\}} \sum_{i=1}^{n} q_{ML}(x_{m,i}, y_i),
\]

where \( x_{m,i} \) denotes the \( i \)-symbol of the \( n \)-th codeword, \( x_m \), and

\[
q_{ML}(x,y) \triangleq \log W(y|x)
\]

(1)

(2)
is referred to as the ML metric.

This can be generalized to a sub-optimal (mismatched) decoding rule [2, Ch. 2]:

\[
\hat{m} = \arg \max_{m \in \{1, \ldots, M_n\}} \sum_{i=1}^{n} q(x_i, m, y_i),
\]

(3)
defined by a single-letter mapping \( q : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{-\infty\} \).

Throughout the paper we assume that ties are broken uniformly between the maximizers. Further we assume that

\[
W(y|x) > 0 \Rightarrow q(x, y) > -\infty
\]

(4)
as otherwise without loss of generality any symbol \( x \) for which there exists \( y \) satisfying \( W(y|x) > 0 \) and \( q(x, y) = -\infty \) should not be used in the codebook, since any codeword containing this symbol results in erroneous decoding with probability \( W(y|x) \), which is bounded away from zero, regardless of the block length.

Denote the random variable corresponding to the decoded message by \( \hat{M}_q(Y) \), and the average error probability as \( P_e(W, C_n, q) = \Pr[\hat{M}_q(Y) \neq M] \).

A rate \( R \) is said to be achievable with decoding metric \( q \) if there exists a sequence of codebooks \( C_n, n = 1, 2, \ldots \) such that \( \frac{1}{n} \log |C_n| \geq R \) and \( \lim_{n \to \infty} P_e(W, C_n, q) = 0 \). The channel capacity w.r.t. metric \( q \), denoted \( C_q(W) \), is defined as the supremum of achievable rates, and is referred to as the mismatch capacity.

Evidently, Shannon’s channel capacity \( C(W) \) can be viewed in fact as the channel capacity w.r.t. the metric \( q_{\text{sl}} \); that is,

\[
C(W) = C_q(W)|_{q=q_{\text{sl}}=\log W}.
\]

(5)

Other quantities that concern channel coding with zero-error and are relevant to our analysis are defined next. The zero-error capacity of channel \( W \), denoted \( C_0(W) \), is defined as the supremum of rates for which there exists a sequence of codebooks \( C_n, n = 1, 2, \ldots \) such that the average error probability with maximum ML decoding is equal to zero. The zero-error capacity with decoding metric \( q \), \( C_{0,q}(W) \), is defined as the supremum of rates for which there exists a sequence of codebooks \( C_n, n = 1, 2, \ldots \) such that \( P_e(W, C_n, q) = 0 \) where ties must be broken uniformly among the minimizers.

A rate-exponent pair \((R, E)\) is said to be achievable for channel \( W \) with decoding metric \( q \) if there exists a sequence of codebooks \( C_n, n = 1, 2, \ldots \) such that for all \( n, \frac{1}{n} \log |C_n| \geq R \) and

\[
\liminf_{n \to \infty} -\frac{1}{n} \log P_e(W, C_n, q) \geq E.
\]

(6)

The reliability function of the channel with decoding metric \( q \) is the supremum of achievable error exponents as a function of the code rate, and is denoted by \( E^q(R, W) \). The reliability function with the optimal ML decoding metric \( q_{\text{sl}}(x, y) = \log W(y|x) \) is denoted \( E(R, W) \); i.e.,

\[
E(R, W) = E(R, W)|_{q=q_{\text{sl}}}. \tag{7}
\]

We next introduce some notation. For a given sequence \( x \in \mathcal{X}^n \), where \( \mathcal{X} \) is a finite alphabet, \( \hat{P}_x \) denotes the empirical distribution on \( \mathcal{X} \) extracted from \( x \); in other words, the vector \( \{\hat{P}_x(x), x \in \mathcal{X}\} \), where \( \hat{P}_x(x) \) is the relative frequency of the symbol \( x \) in the vector \( x \). The type-class of \( x \) is the set of \( x' \in \mathcal{X}^n \) such that \( \hat{P}_{x'} = \hat{P}_x \), which is denoted \( T_x(\hat{P}_x) \). The set of all probability distributions on \( \mathcal{X} \) is denoted by \( \mathcal{P}(\mathcal{X}) \), and the set of empirical distributions of order \( n \) on alphabet \( \mathcal{X} \) is denoted \( \mathcal{P}_n(\mathcal{X}) \).

For \( P \in \mathcal{P}_n(\mathcal{X}) \), we say that the codebook \( C_n = \{x_i\}_{i=1}^{M_n} \) is a \( P \)-constant composition codebook if for all \( i \) satisfies \( \hat{P}_{x_i} = P \). Define the highest achievable exponent with \( P \)-constant composition codebooks of block length \( n \) and \( q \)-decoding as

\[
e^q_n(R, P, W) \triangleq \max_{C_n \subseteq \mathcal{P}_n(P)} \frac{1}{n} \log P_e(q, W, C_n). \tag{8}
\]

Using standard arguments that follow from the fact that the type of number-classes grows polynomially with \( n \), it can be shown\(^1\) that

\[
E^q(R, W) = \liminf_{n \to \infty} \max_{P \in \mathcal{P}(\mathcal{X})} e^q_n(R, P, W). \tag{9}
\]

III. PREVIOUS BOUNDS AND RELATED RESULTS

In this section we mention previous bounds and results that are most relevant to this work. For a comprehensive survey the reader is referred to the additional aforementioned works and references therein.

The classical sphere-packing upper bound on \( E(R, W) \) is given by:

\[
E_{sp}(R, W) = \max_{P \in \mathcal{P}(\mathcal{X})} E_{sp}(R, P, W) \tag{10}
\]

where \( \mathcal{P}(\mathcal{X}) \) denotes the set of all probability distributions on \( \mathcal{X} \) and

\[
E_{sp}(R, P, W) = \min_{P_{Y|X} : I(P \times P_{Y|X}) \leq R} D(P_{Y|X} \| W|P), \tag{11}
\]

which holds for the ML metric, and thus also for any other metric. The expurgated lower bound on the reliability function for coding with decoding metric \( q \), \( E^q(R, W) \) ([1], [11], [18]), is given by:

\[
E^q_{lx}(R, W) = \max_{P \in \mathcal{P}(\mathcal{X})} \min_{V : \text{VX = VX = } P, \text{ Eq(X,Y) \geq Eq(X,Y), I(X;Y) \leq R}} \left[ D(V_{Y|X} \| W|P) + I(\tilde{X};Y, X) - R. \right. \tag{12}
\]

Shannon, Gallager and Berlekamp [3] showed that for \( q = q_{\text{sl}} \), the bound is tight, that is,

\[
E(0^+, W) = E_{lx}(0^+, W) \triangleq E^q_{lx}(0^+, W)|_{q=q_{\text{sl}}}, \tag{13}
\]

for channels satisfying \( C_0(W) = 0 \). Another important result of [3] is the derivation of the following combined upper bound

\[
E_{sl-xp}(R, W) = \tag{14}
\]

\(^1\)Any codebook \( C_n \subseteq \mathcal{X}^n \) of size \( e^{nR} \) contains at least one constant-composition subset \( C'_n \) whose size satisfies \( \frac{1}{n} \log |C'_n| \geq R - \frac{\log(n+1)}{n} \).
\[
\left\{ \begin{array}{ll}
-\frac{\mathbb{E}(0^+, W) - E_{sp}(R_W^*, W)}{R_W^*} R + E(0^+, W) & R < R_W^* \\
E_{sp}(R, W) & R > R_W^*
\end{array} \right.
\]  

(14)

where \( R_W^* \) signifies the \( R \)-coordinate of the point at which the tangential line from the point \((0, E(0^+, W))\) meets the \( E_{sp}(\cdot, W) \) curve in the \((R, E)\)-plane. The linear part of the curve; i.e., for \( R \in (0, R_W^*) \), is referred to as the straight-line bound.

We recently obtained results [17] which provide the mismatch counterpart of the sphere-packing bound:

\[
E_{sp}^q(R, P, W) \triangleq \min_{P_{Y|X} \in W_q(P)} D(P_{Y|X} || W|P),
\]

(15)

where \( W_q(P_X) \) stands for the following set of channels:

\[
\left\{ P_{Y|X} : \begin{array}{c}
V_{U,X,Z,Y}|V_{X,Y} = p_{X,Y|Z} \\
V_{U,X,Z} = V_{X,Z} \\
\tilde{X} - (U,Z) - X \\
(U,U) - (X,Z) - Y
\end{array} \right\}
\]

and where \( U \) is an auxiliary random variable with alphabet \(|U| \leq |X|^2|Z|\), and the condition \( V_{U,X,Z} = V_{U,X,Z,Y} \) signifies that for all \((u, x_1, x_2, z)\), \( V_{U,X,Z}(u, x_1, x_2, z) = V_{U,X,Z}(u, x_1, x_2, z)\).

In [15], the expurgated bound \( E_{sp}^q(R, W) \) was shown to be tight at \( R = 0^+ \) also in the mismatched case for the wide class of balanced channel-metric pairs defined as follows:

**Definition 1.** ([15]) A discrete memoryless channel \( W_{Y|X} \) and a decoding metric \( q \) form a balanced pair if \( C_{0,q}(W) = 0 \) and

\[
\max_{y: W(y|x) > 0} [q(x, y) - q(\tilde{x}, y)] = \min_{y: W(y|x) > 0} [q(x, y) - q(\tilde{x}, y)]
\]

\[
\max_{y: W(y|x) > 0, q(x,y) \neq -\infty, q(\tilde{x},y) \neq -\infty} [q(x, y) - q(\tilde{x}, y)]
\]

\[
\min_{y: W(y|x) > 0, q(x,y) \neq -\infty, q(\tilde{x},y) \neq -\infty} [q(x, y) - q(\tilde{x}, y)]
\]

(16)

As noted in [15], all channels and decoding metrics such that \( C_{0,q}(W) = 0 \) and

\[
W(y|x) > 0 \iff q(x, y) > -\infty
\]

are balanced, and obviously form a very important wide class of channel-metric pairs, which includes the ML metric case of channels for which \( C_0(W) = 0 \).

**IV. MAIN RESULTS**

This section is devoted to presenting the new upper bound on \( E^n(R, W) \) and related results.

**A. The Main Theorem**

Let \( W_{Y|Z|X} \) be a DMC from \( X \) to \( Y \times Z \). Throughout this paper, we adopt the shorthand notation that \( W \) without subscript signifies the original single-user channel \( W_{Y|X} \). Whenever we refer to other marginal distributions; i.e., \( W_{Y|Z}, W_{Z|X}, W_{Z|Y} \) or \( W_{Y|Z|X} \), the subscript is mentioned explicitly.

Let \( U \) denote a random variable over alphabet \( \mathcal{U} \) of cardinality \( |\mathcal{U}| \leq |X|^2|Z| + 1 \). Consider the following function

\[
E^q(R, P, W_{Y|Z|X}) \triangleq \min_{P_{X}} \mathbb{E}^q(P_{XZ}, W_{Y|Z|X}),
\]

(18)

\[
\mathbb{E}(R, P, W_{Y|Z|X}) \triangleq \max_{P_{X|U}X,Z,Y} \min_{P_{Y|X,Z}} g(P_{XZU\tilde{X}Y}, W_{Y|Z|X})
\]

(19)

and \( g(P_{XZU\tilde{X}Y}, W_{Y|Z|X}) = D(P_{Y|X} || W_{Y|Z|X}|P) + I(\tilde{X}; Y|X, Z) \). For the ML decoding metric denote

\[
\overline{E}(R, P, W_{Y|Z|X}) = \mathbb{E}^q(R, P, W_{Y|Z|X})_{q=ml}.
\]

(20)

Due to (9), our first result is presented in terms of an upper bound on \( e_q^n(R, P, W) \). The proofs of the results in this paper can be found in [19].

**Theorem 1.** If \( (q, W_{Y|X}) \) is balanced, then for any \(|Z| < \infty, \epsilon > 0 \) and \( n \) sufficiently large, it holds that for any \( W_{Z|XY} \) and any \( P \in \mathcal{P}_n(X) \),

\[
e_q^n(R, P, W_{Y|X}|X) \leq \mathbb{E}^q(R - \epsilon, P, W_{Y|X}|X) + \epsilon.
\]

(21)

Note that \( W_{Y|Z|X} = W_{Y|X} \times W_{Z|XY} \), and that any choice of \( W_{Z|XY} \) in (21) is valid, and can be optimized as a function of \((P, R, q, W_{Y|X}, \epsilon, n)\). In the proof of Theorem 2 to follow, we show that there exists a certain choice of \( W_{Z|XY} \) which results in a bound that is at least as tight as several known bounds.

**Proof Outline:** As mentioned above, the bound relies on the approach of multicast transmission with an auxiliary receiver. The idea behind this proof technique is very simple (Fig. 1).

We extend the single-user channel from \( X \) to \( Y \) to have an additional output \( Z \). The \( Z \)-receiver, serves as a genie to the \( Y \)-receiver, and provides it with the list of all the codewords, which lie in the same conditional type-class given the received signal \( Z \) as that of the true transmitted one, including the latter. To accomplish this, another genie informs the \( Z \)-receiver of the actual joint distribution of \( Z \) and the true codeword \( X \). Thus, the \( Z \)-receiver can be viewed as a “genie-aided-genie”.
The $Y$-receiver needs only to search within the narrowed list for the codeword that maximizes the decoding metric (either maximum likelihood or mismatched) rather than within the entire codebook, and therefore, analyzing this setup yields an upper bound on the exponent of the error probability.

We further show that provided that the list size is large enough, most of the list can be partitioned into disjoint subsets (sub-lists) of large size, where within each sub-list, the pairwise joint distribution given $z$ is approximately symmetric and approximately the same for all pairs. This extends a result of [6], which can be viewed as the existence of such single sub-list for the case where $z$ is null.

This above partitioning of the list enables to show that if one considers channels with $z$ output for which with high probability the list size is large enough, then the average error probability conditioned on $z$ is essentially lower bounded by the average pairwise error probability of codewords within a sub-list. The latter can be lower bounded using Plotkin’s counting trick as in [3] and the above mentioned Blinkovsky’s idea [7] of using Ramsey-theoretic result by Komló [8] similar to [6], [15].

B. A Dual Form of the Bound

Next, we present a semi-dual form of the bound $E^q(R, P, W_{YZ} | X)$, whose proof appears in [19].

Let

$$
\zeta_{x, \bar{z}}(s, W_{YZ} | X) = \log \sum_y W(y|x, z) e^{\eta_q(P_{XZU}, W_{Y|X})}
$$

with cases of a strict inequality for $\alpha \in (0, 1)$. Further, if $(q, W)$ is a balanced pair then

$$
E^q(0^+, W) = \max_{P \in P(X)} E^q_0(P, W) = E^q_{ex}(0^+, W), \quad (27)
$$

$$
\max_{P \in P(X)} E^q(R, P, W_{YZ} | X) \bigg|_{q = \log W} \leq E_{sp}(R, W), \quad (28)
$$

The proof of Theorem 2 can be found in [19], and as mentioned above, the equality $E^q(0^+, W) = E^q_{ex}(0^+, W)$ was proved in [15].

Next, it is easy to verify (see [19]) that our bound is at least as tight as our previous best known upper bound (see (15)).

**Proposition 2.**

$$
\inf_{W_{YZ} | X} E^q(R, P, W_{YZ} | X) \leq E^q_{sp}(R, P, W). \quad (29)
$$

Recall the definitions (22)-(23). In the next proposition we present a looser bound that takes a simpler form for both the matched and mismatched metric cases. Denote

$$
E^q_B(R, P, W) \triangleq \min_{P_{XZ} : P_X = P, P_z | X \approx R} \max_{q \in P_{V \approx Z}} \eta_q(P_{XZU}, W_{Y|X}) \quad (30)
$$

$$
E^q_{sp}(R, P, W) \triangleq \min_{P_{XZ} : P_X = P, P_z | X \approx R} \max_{q \in P_{V \approx Z}} \sum_{u, x, z} P_{XZU}(x, z, u) \times P_{X|U}(\bar{z}|u, z) \log \frac{1}{\sum_y \sqrt{W(y|x)W(y|\bar{x})}}. \quad (31)
$$

We have the following result:

**Proposition 3.** The following inequality holds

$$
\inf_{W_{YZ} | X} E^q(R, P, W_{YZ} | X) \leq E^q_B(R, P, W), \quad (32)
$$

and in particular

$$
\inf_{W_{YZ} | X} E^q(R, P, W_{YZ} | X) \leq E^q_{sp}(R, P, W), \quad (33)
$$

with cases of strict inequality.

The proof appears in [19], where we also outline an alternative proof for the looser bound $E(R, P, W) \leq E^q_{sp}(R, P, W)$ that is based mostly on the derivation of the bound [6, Eq. (89)], applied to the case $L = 1$. It is interesting to note that in this alternative proof of the looser bound, the output signal $Z$ is substituted by the deterministic sequence whose existence is guaranteed by a covering argument (see [6, Lemma 5], and [5, Lemma 1]).

D. A Looser Bound and an Example: the BSC with ML Decoder

It is desirable to evaluate the bound $E^q(R, P, W_{YZ} | X)$ or simplify it, and in particular, to tackle the maximization over $P_{U | XZ}$. In the case of the BSC, it can be shown (see details in [19]) that the following bound holds.

**Lemma 1.** For the BSC $W_{Y | X}$ with ML decoding metric it holds that

$$
E^q(R, P, W_{YZ} | X)
$$
\[
\min_{\gamma: 1 - h_2(\gamma) \leq R} \frac{1}{2} \left[ \gamma \log \frac{\gamma^2}{W_{Z|X}(0|0)W_{Z|X}(0|1)} \right] + \gamma \log \frac{W_{Z|X}(0|0)W_{Z|X}(1|1)}{\gamma^2} \\
+ \frac{1}{2} \sup_{s \geq 0} \left\{ \sum_{z: d_{z,s,q,W_{Y|X}}(0,1) > 0} d_{z,s,q,W_{Y|X}}(0,1) \right\} + \sum_{z: d_{z,s,q,W_{Y|X}}(1,0) > 0} d_{z,s,q,W_{Y|X}}(1,0),
\]

where \( h_2(x) = -x \log(x) - (1 - x) \log(1 - x) \) is the binary entropy function,

\[
d_{z,s,q,W_{Y|X}}(x, \bar{x}) = \frac{1}{2} \left( \zeta_{x,\bar{x},z}(W_{Y|X}) + \zeta_{x,\bar{x},z}(W_{Y|X}) \right).
\]

and \( \zeta_{x,\bar{x},z}(s,W_{Y|X}) \) is defined in (22).

A numerical calculation of the bound (34), denoted \( \mathcal{E}_{sym} \), for \(|Z| = 2\) is depicted in Fig. 2 which is calculated using a grid search for the optimal choice of \( W_{Z|X,Y} \), however, since any value of \( W_{Z|X,Y} \) yields a valid upper bound, this implies that the actual bound (34) may be tighter than the one depicted. Moreover, the bound \( \inf_{W_{Z|X,Y}} \mathcal{E}_{B}^2(R, P, W_{Y|X}) \) may be even tighter, and in addition \(|Z|\) can be taken larger than 2. For the sake of comparison, the upper bounds \( E_{sp}(R) \), and the amended Blahut bound \( E_B(R) \) are also depicted, as well as the straight line bound, and the lower bound:

\[
E_{LB}(R) = \begin{cases} 
-\delta_{GV}(R) \cdot \log \left( \sqrt{2p(1-p)} \right) & R \leq R_{min} \\
1 - \log \left( 1 + 2\sqrt{2p(1-p)} \right) - R & R \in [R_{min}, R_{crit}] \\
E_{sp}(R) & R \geq R_{crit}
\end{cases}
\]

where \( \delta_{GV}(R) \) is the solution to the equation \( 1 - h_2(\delta) = R \), and \( R_{min} = 1 - h_2 \left( \frac{\sqrt{2p(1-p)}}{1+2\sqrt{2p(1-p)}} \right) \). Exploiting symmetries as well as other special properties, tighter bounds compared to the general DMC case have been derived for the BSC, which are not depicted in this figure (see [9]), and seem tighter compared to our numerical calculation. In order to compare \( \inf_{W_{Z|X,Y}} \mathcal{E}_{B}^2(R, P, W_{Y|X}) \) with larger \(|Z|\) (rather than the looser version (34) with \(|Z| = 1\)) applied to the BSC case to the bound of [9], further study is needed. Nevertheless, the BSC serves as an example that shows that our bound strictly improves on the straight line bound as well as the Blahut amended bound and the sphere packing bound.

**References**

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

[2] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems.* Cambridge University Press, 2011.

[3] C. Shannon, R. Gallager, and E. Berlekamp, “Lower bounds to error probability for coding on discrete memoryless channels. II,” *Information and Control*, vol. 10, no. 5, pp. 522–552, 1967. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0019995867901200

[4] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, “Lower bounds to error probability for coding on discrete memoryless channels. I,” *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967.

[5] R. Blahut, “Composition bounds for channel block codes,” *IEEE Trans. Inf. Theory*, vol. 23, no. 6, pp. 656–674, 1977.

[6] M. Bondaschi and M. Dalai, “A revisitation of low-rate bounds on the reliability function of discrete memoryless channels for list decoding,” *IEEE Transactions on Information Theory*, vol. 68, no. 5, pp. 3282–2838, 2022.

[7] V. M. Blinovskiy, “New approach to estimation of the decoding error probability,” *Problems of information transmission*, vol. 38, no. 1, pp. 16–19, 2002.

[8] J. Komlós, “A strange pigeon-hole principle,” *Order*, vol. 7, no. 2, pp. 107–113, 1990.

[9] M. V. Burnashev, “On reliability function of BSC: Expanding the region, where it is known exactly,” *Problems of Information Transmission*, vol. 51, no. 4, pp. 307–325, 2015.

[10] E. A. Haroutunian, M. E. Haroutunian, and A. N. Harutyunyan, “Reliability criteria in information theory and statistical hypothesis testing,” *Foundations and Trends® in Communications and Information Theory*, vol. 4, no. 2–3, pp. 97–263, 2008. [Online]. Available: http://dx.doi.org/10.1561/0100000008

[11] I. Csiszár and J. Körner, “Graph decomposition: A new key to coding theorems,” *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, 1981.

[12] A. Somekh-Baruch, “On achievable rates and error exponents for channels with mismatched decoding,” *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 727–740, Feb 2015.

[13] J. Scarlett, L. Peng, N. Merhav, A. Martínez, and A. Guillén i Fábregas, “Expurgated random-coding ensembles: Exponents, refinements, and connections,” *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4449–4462, Aug. 2014.

[14] J. Scarlett, A. Guillén i Fábregas, A. Somekh-Baruch, and A. Martínez, “Information-theoretic foundations of mismatched decoding,” *Foundations and Trends® in Communications and Information Theory*, vol. 16, no. 5–6, pp. 577–644, Sept. 2020.
[15] M. Bondaschi, A. Guillén i Fàbregas, and M. Dalai, “Mismatched decoding reliability function at zero rate,” IEEE Transactions on Information Theory, pp. 1–1, 2021.

[16] E. A. Kangarshahi and A. Guillén i Fàbregas, “A sphere-packing error exponent for mismatched decoding,” arXiv:2111.04797 (2021).

[17] A. Somekh-Baruch, “Upper bounds on the mismatched reliability function and capacity using a genie receiver,” 2022. [Online]. Available: https://arxiv.org/abs/2203.08524, to appear in IEEE Transactions on Information Theory.

[18] A. Somekh-Baruch, J. Scarlett, and A. Guillén i Fàbregas, “Generalized random Gilbert-Varshamov codes,” IEEE Transactions on Information Theory, vol. 65, no. 6, pp. 3452–3469, 2019.

[19] A. Somekh-Baruch, “An upper bound on the reliability function of the DMC,” 2022. [Online]. Available: https://arxiv.org/abs/2209.00968