# On the Bee-Identification Error Exponent with Absentee Bees 

Anshoo Tandon, Member, IEEE, Vincent Y. F. Tan, Senior Member, IEEE, and Lav R. Varshney, Senior Member, IEEE


#### Abstract

The "bee-identification problem" was formally defined by Tandon, Tan and Varshney [IEEE Trans. Commun. (2019) [Online early access]], and the error exponent was studied. This work extends the results for the "absentee bees" scenario, where a small fraction of the bees are absent in the beehive image used for identification. For this setting, we present an exact characterization of the bee-identification error exponent, and show that independent barcode decoding is optimal, i.e., joint decoding of the bee barcodes does not result in a better error exponent relative to independent decoding of each noisy barcode. This is in contrast to the result without absentee bees, where joint barcode decoding results in a significantly higher error exponent than independent barcode decoding. We also define and characterize the "capacity" for the bee-identification problem with absentee bees, and prove the strong converse for the same.


Index Terms-Bee-identification problem, absentee bees, noisy channel, error exponent, capacity, strong converse.

## I. Introduction

The problem of bee-identification with absentee bees can be described as follows. Consider a group of $m$ different bees, in which each bee is tagged with a unique barcode for identification purposes in order to understand interaction patterns in honeybee social networks [1], [2]. Assume a camera takes a picture of the beehive to study the interactions among bees. The beehive image output (see Fig. 1) can be considered as a noisy and unordered set of barcodes. In this work, we consider the "absentee bees" scenario, in which some bee barcodes are missing in the image used to decode the identities of the bees. This scenario can arise, for instance, when some of the bees fly away from the beehive, or when some of the bees (or their barcodes) are occluded from view. Posing as an information-theoretic problem, we quantify the error probability of identifying the bees still present in the finiteresolution beehive image through the corresponding (largest or best) error exponent.

The barcode for each bee is represented as a binary vector of length $n$, and the bee barcodes are collected in a codebook

This work was supported in part by a Singapore Ministry of Education Tier 2 grant (R-263-000-C83-112).
A. Tandon is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583 (email: anshoo.tandon@gmail.com).
V. Y. F. Tan is with the Department of Electrical and Computer Engineering, and with the Department of Mathematics, National University of Singapore, Singapore (email: vtan@nus.edu.sg).
L. R. Varshney is with the Coordinated Science Laboratory and the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (email: varshney @illinois.edu)


Fig. 1: Bees tagged with barcodes (photograph provided by T. Gernat and G. Robinson).
$\mathcal{C}$ comprising $m$ rows and $n$ columns, with each row corresponding to a bee barcode. As shown in Fig. 2, the channel first permutes the $m$ rows of $\mathcal{C}$ with a random permutation $\pi$ to produce $\mathcal{C}_{\pi}$, where the $i$-th row of $\mathcal{C}_{\pi}$ corresponds to the $\pi(i)$-th row of $\mathcal{C}$. Next, the channel deletes $k$ rows of $\mathcal{C}_{\pi}$, to model the scenario in which $k$ bees, out of a total of $m$ bees, are absent in the beehive image. Without loss of generality, we assume that the channel deletes the last $k$ rows of $\mathcal{C}_{\pi}$ to produce $\mathcal{C}_{\pi_{(m-k)}}$, where $\pi_{(m-k)}$ denotes an injective mapping from $\{1, \ldots, m-k\}$ to $\{1, \ldots, m\}$ and corresponds to the restriction of permutation $\pi$ to only its first $m-k$ entries. Finally, the channel adds noise, modeled as a binary symmetric channel (BSC) with crossover probability $p$ with $0<p<0.5$, to produce $\tilde{\mathcal{C}}_{\pi_{(m-k)}}$ at the channel output. We assume the decoder has knowledge of codebook $\mathcal{C}$, and its task is to recover the channel-induced mapping $\pi_{(m-k)}$ using the channel output $\tilde{\mathcal{C}}_{\pi_{(m-k)}}$. Note that $\pi_{(m-k)}$ directly ascertains the identity of all $m-k$ bees present in the image.
When $j=\pi(i)$ and the $j$-th row of codebook $\mathcal{C}$ is denoted $\boldsymbol{c}_{j}=\left[\begin{array}{llll}c_{j, 1} & c_{j, 2} & \cdots & c_{j, n}\end{array}\right]$, then the $i$-th row of $\mathcal{C}_{\pi}$ is equal to $\boldsymbol{c}_{j}$. For $1 \leq i \leq m-k$, the $i$-th row of $\tilde{\mathcal{C}}_{\pi_{(m-k)}}$, denoted $\tilde{\boldsymbol{c}}_{i}$, is a noisy version of $\boldsymbol{c}_{j}=\boldsymbol{c}_{\pi(i)}$ and we have

$$
\begin{align*}
\operatorname{Pr}\left\{\tilde{\boldsymbol{c}}_{i} \mid \boldsymbol{c}_{\pi(i)}\right\} & =p^{d_{i}}(1-p)^{n-d_{i}}, \quad 1 \leq i \leq m-k \\
\operatorname{Pr}\left\{\tilde{\mathcal{C}}_{\pi_{(m-k)}} \mid \mathcal{C}, \pi_{(m-k)}\right\} & =\prod_{i=1}^{m-k} \operatorname{Pr}\left\{\tilde{\boldsymbol{c}}_{i} \mid \boldsymbol{c}_{\pi(i)}\right\} \\
& =\prod_{i=1}^{m-k} p^{d_{i}}(1-p)^{n-d_{i}} \tag{1}
\end{align*}
$$

## Effective Channel



Fig. 2: Effective channel for the bee-identification problem with $k$ absentee bees.
where $d_{i} \triangleq \mathrm{~d}_{\mathrm{H}}\left(\tilde{\boldsymbol{c}}_{i}, \boldsymbol{c}_{\pi(i)}\right)$ denotes the Hamming distance between vectors $\tilde{\boldsymbol{c}}_{i}$ and $\boldsymbol{c}_{\boldsymbol{\pi}(i)}$.

We remark that the bee-identification problem formulation has other applications in engineering, such as packagedistribution to recipients from a batch of deliveries with noisy address labels, and identification of warehouse products using wide-area sensors [1]. In a related work on identification via permutation recovery [3], the identification of the respective distributions of a set of observed sequences (in which each sequence is generated i.i.d. by a distinct distribution) was analyzed. Other models and applications, in which permutation recovery arises naturally, are discussed in [4].

In another related work, the fundamental limit of data storage via unordered DNA molecules was investigated in [5], while the corresponding capacity results were extended to a noisy setting in [6]. Here, we remark that the effective channel in [6] is closely related to the bee-identification channel. Data communication over permutation channels with impairments, such as insertions, deletions, and substitutions, was analyzed in [7]. Note that the effective channel for the bee-identification problem differs from the communication channel in [7] in two aspects: (i) The input to the channel in the bee-identification problem is the entire codebook (see Fig. 2), instead of just a codeword belonging to the codebook; (ii) The channel in Fig. 2 only permutes the rows of the codebook, but does not permute the letters within a row.

## A. Absentee Bee-Identification Error Exponent

The bee-identification problem involves designing a decoder to detect the channel-induced mapping $\pi_{(m-k)}$ by using the knowledge of codebook $\mathcal{C}$ and channel output $\tilde{\mathcal{C}}_{\pi_{(m-k)}}$. The decoder corresponds to a function $\phi$ that takes $\tilde{\mathcal{C}}_{\pi_{(m-k)}}$ as an input and produces a map $\nu:\{1, \ldots, m-k\} \rightarrow\{1, \ldots, m\}$ where $\nu(i)$ corresponds to the index of the transmitted codeword which produced the received word $\tilde{\boldsymbol{c}}_{i}$, where $1 \leq i \leq$ $m-k$. The indicator variable for the bee-identification error is defined as

$$
\mathcal{D}\left(\nu, \pi_{(m-k)}\right) \triangleq \begin{cases}1, & \text { if } \nu \neq \pi_{(m-k)} \\ 0, & \text { if } \nu=\pi_{(m-k)}\end{cases}
$$

Let $\Upsilon$ denote the set of all injective maps from $\{1, \ldots, m-$ $k\}$ to $\{1, \ldots, m\}$. For a given codebook $\mathcal{C}$ and decoding function $\phi$, the expected bee-identification error probability over the $\operatorname{BSC}(p)$ is

$$
\begin{equation*}
D(\mathcal{C}, p, k, \phi) \triangleq \mathbb{E}_{\pi_{(m-k)}}\left[\mathbb{E}\left[\mathcal{D}\left(\nu, \pi_{(m-k)}\right)\right]\right] \tag{2}
\end{equation*}
$$

where the inner expectation is over the distribution of $\tilde{\mathcal{C}}_{\pi_{(m-k)}}$ given $\mathcal{C}$ and $\pi_{(m-k)}$ (see (1)), and the outer expectation is over
the uniform distribution of $\pi_{(m-k)}$ over $\Upsilon$. Note that (2) can be equivalently expressed as

$$
\begin{equation*}
D(\mathcal{C}, p, k, \phi)=\mathbb{E}_{\pi_{(m-k)}}\left[\operatorname{Pr}\left\{\nu \neq \pi_{(m-k)}\right\}\right] . \tag{3}
\end{equation*}
$$

Let $\mathscr{C}(n, m)$ denote the set of all binary codebooks of size $m \times n$, i.e. binary codebooks with $m$ codewords, each having length $n$. Now, for given values of $n, m$, and $k$, define the minimum expected bee-identification error probability as

$$
\begin{equation*}
\underline{D}(n, m, p, k) \triangleq \min _{\mathcal{C}, \phi} D(\mathcal{C}, p, k, \phi) \tag{4}
\end{equation*}
$$

where the minimum is over all codebooks $\mathcal{C} \in \mathscr{C}(n, m)$, and all decoding functions $\phi$. The exponent corresponding to the minimum expected bee-identification error probability is given by $-\frac{1}{n} \log \underline{D}(n, m, p, k)$. Note that we take all logarithms to base 2 , unless stated otherwise.

## B. Our Contributions

We consider the bee-identification problem with a constant fraction of "absentee bees", and provide an exact characterization of the corresponding error exponent; this is done via Theorem 1 We show that joint decoding of the bee barcodes does not result in a better error exponent relative to the independent decoding of each noisy barcode. This is in contrast to the result without absentee bees [1], where joint barcode decoding results in a significantly higher error exponent than independent barcode decoding.

Secondly, we define and characterize the "capacity" (i.e., the supremum of all code rates for which the error probability can be driven to 0 ) of the bee-identification problem with absentee bees via Theorem 2 Further, we prove the strong converse showing that for rates greater than the capacity, the error probability tends to 1 as the blocklength (length of barcodes) tends to infinity.

Lastly, via Theorem 3, we show that for low rates, the error exponent when the fraction of absentee bees tends to zero, is strictly lower than the corresponding exponent for the case without absentee bees. This implies a discontinuity in the error exponent function when the fraction of absentee bees $\alpha$ tends to 0 , highlighting the dichotomy of its behavior when $\alpha>0$ and when $\alpha=0$. On the one hand, independent barcode decoding is optimal even when arbitrarily small fraction of bees are absent, whereas on the other hand, joint barcode decoding provides higher exponent when no bees are absent.

## II. Bounds on the Error Probability

In this section, we present finite-length bounds on the minimum expected bee-identification error probability, $\underline{D}(n, m, p, k)$. The upper bound on $\underline{D}(n, m, p, k)$ is presented
in Section II-A using a naïve decoding strategy in which each noisy barcode is decoded independently, while the lower bound on $\underline{D}(n, m, p, k)$ is presented in Section II-B using joint maximum likelihood (ML) decoding of barcodes.

## A. Independent decoding based upper bound on $\underline{D}(n, m, p, k)$

We present an upper bound on $\underline{D}(n, m, p, k)$ based on two ideas: (i) independent decoding of each barcode, and (ii) the union bound. Independent barcode decoding is a naïve strategy where, for $1 \leq i \leq m-k$, the decoder picks $\tilde{\boldsymbol{c}}_{i}$, the $i$-th row of $\tilde{\mathcal{C}}_{\pi_{(m-k)}}$, and then decodes it to $\nu(i)=\arg \min _{1 \leq j \leq m} \mathrm{~d}_{\mathrm{H}}\left(\tilde{\boldsymbol{c}}_{i}, \boldsymbol{c}_{j}\right)$. If there is more than one codeword at the same minimum Hamming distance from $\tilde{\boldsymbol{c}}_{i}$, then any one of the corresponding codeword indices is chosen uniformly at random.

We denote the decoding function $\phi$ corresponding to independent barcode decoding as $\phi_{\mathrm{I}}$. Then, for a given codebook $\mathcal{C}$, it follows from (3) that

$$
\begin{align*}
D\left(\mathcal{C}, p, k, \phi_{\mathrm{I}}\right) & =\mathbb{E}_{\pi_{(m-k)}}\left[\operatorname{Pr}\left\{\nu \neq \pi_{(m-k)}\right\}\right] \\
& \leq \sum_{i=1}^{m-k} \mathbb{E}_{\pi_{(m-k)}}\left[\operatorname{Pr}\left\{\nu(i) \neq \pi_{(m-k)}(i)\right\}\right] \tag{5}
\end{align*}
$$

where the inequality follows from the union bound and the linearity of the expectation operator.

For a scenario in which $m$ binary codewords, each having blocklength $n$, are used for transmission of information over a binary symmetric channel $\operatorname{BSC}(p)$, let $P_{\mathrm{e}}(n, m, p)$ denote the minimum achievable average error probability, where the minimization is over all codebooks $\mathcal{C} \in \mathscr{C}(n, m)$. The following lemma applies (4) and (5) to present an upper bound on $\underline{D}(n, m, p, k)$ in terms of $P_{\mathrm{e}}(n, m, p)$.
Lemma 1. Using independent barcode decoding, the beeidentification error probability $\underline{D}(n, m, p, k)$ can be upper bounded as follows

$$
\begin{equation*}
\underline{D}(n, m, p, k) \leq \min \left\{1,(m-k) P_{\mathrm{e}}(n, m, p)\right\} . \tag{6}
\end{equation*}
$$

Proof: See Appendix A

## B. Joint decoding based lower bound on $\underline{D}(n, m, p, k)$

Recall $\Upsilon$ denotes the set of all injective maps from $\{1, \ldots, m-k\}$ to $\{1, \ldots, m\}$. With joint ML decoding of barcodes using a given codebook $\mathcal{C}$, the decoding function $\phi$ takes the channel output $\tilde{\mathcal{C}}_{\pi_{(m-k)}}$ as an input, and produces the map

$$
\begin{equation*}
\nu=\underset{\sigma \in \Upsilon}{\arg \min } \mathrm{d}_{\mathrm{H}}\left(\tilde{\mathcal{C}}_{\pi_{(m-k)}}, \mathcal{C}_{\sigma}\right) \tag{7}
\end{equation*}
$$

where $\mathcal{C}_{\sigma}$ denotes a matrix with $m-k$ rows and $n$ columns whose $i$-th row is equal to the $\sigma(i)$-th row of $\mathcal{C}$, and $\mathrm{d}_{\mathrm{H}}\left(\tilde{\mathcal{C}}_{\pi_{(m-k)}}, \mathcal{C}_{\sigma}\right) \triangleq \mid\left\{(i, j): \tilde{\mathcal{C}}_{\pi_{(m-k)}}(i, j) \neq \mathcal{C}_{\sigma}(i, j), 1 \leq i \leq\right.$ $m-k, 1 \leq j \leq n\} \mid$. For this joint ML decoding scheme, we denote the decoding function as $\phi_{\mathrm{J}}$. As $\pi_{(m-k)}$ is uniformly distributed over $\Upsilon$, the joint ML decoder minimizes the error probability [8, Thm. 8.1.1], and from (4) we have

$$
\begin{equation*}
\underline{D}(n, m, p, k)=\min _{\mathcal{C} \in \mathscr{C}(n, m)} D\left(\mathcal{C}, p, k, \phi_{\mathrm{J}}\right) . \tag{8}
\end{equation*}
$$

The following lemma uses (8) to present a lower bound on $\underline{D}(n, m, p, k)$ in terms of $P_{\mathrm{e}}(n, k+1, p)$.

Lemma 2. Let $0<\varepsilon<1 / 2$, and let $k>1 / \varepsilon$. Then, the beeidentification error probability $\underline{D}(n, m, p, k)$ using joint $M L$ decoding of barcodes can be lower bounded as follows

$$
\begin{equation*}
\underline{D}(n, m, p, k)>\frac{1-2 \varepsilon}{2} \min \left\{1,(m-k) \varepsilon P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)\right\} . \tag{9}
\end{equation*}
$$

Furthermore, the error probability $\underline{D}(n, m, p, k)$ can alternatively be lower bounded as follows

$$
\begin{align*}
& \underline{D}(n, m, p, k) \\
& \quad>(1-2 \varepsilon)\left[1-\exp \left(-(m-k) \varepsilon P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)\right)\right] . \tag{10}
\end{align*}
$$

Proof: See Appendix B.
The lower bound in 9 will be used to prove the converse part in Theorem 1 on characterizing the error exponent. On the other hand, the lower bound in (10) helps us to characterize the "capacity" of the bee-identification problem in Theorem 2 and to prove the strong converse for the same problem.

## III. Bee-Identification Exponent and the Optimality of Independent Decoding

In this section, we analyze the exponent of the minimum expected bee-identification error probability, $-\frac{1}{n} \log \underline{D}(n, m, p, k)$. We first present some notation for the bee-identification exponent. Recall that $P_{\mathrm{e}}(n, m, p)$ denotes the minimum achievable average error probability when $m$ binary codewords, each having blocklength $n$, are used for transmission of information over $\operatorname{BSC}(p)$. For a given $R>0$ and $m=\left\lceil 2^{n R}\right\rceil \prod^{1}$ the reliability function of the channel $\operatorname{BSC}(p)$ is defined as follows [9] ${ }^{2}$

$$
\begin{equation*}
E(R, p) \triangleq \limsup _{n \rightarrow \infty}-\frac{1}{n} \log P_{\mathrm{e}}\left(n, 2^{n R}, p\right) \tag{11}
\end{equation*}
$$

Let $\left(R_{n}\right)_{n \in \mathbb{N}}$ be a sequence that converges to $R$, and for a fixed $n$ we define

$$
\begin{equation*}
E\left(n, R_{n}, p\right) \triangleq-\frac{1}{n} \log P_{\mathrm{e}}\left(n, 2^{n R_{n}}, p\right) \tag{12}
\end{equation*}
$$

We will relate $E\left(n, R_{n}, p\right)$ to $E(R, p)$ via Lemma 3 However, in order to establish Lemma 3. we need the fact that $E(R, p)$ is continuous for $R>0$. We remark that although the continuity of $E(R, p)$ (or $E(R, W)$ for a general discrete memoryless channel $W$ ), has been discussed in previous literature (see, e.g., [12, Lem. 1], [13, p. 113], [14, Prop. 8]), a clear and comprehensive proof appears to be elusive. Note that the scenario where the rate is less than the critical rate is of particular interest, because it is well known that $E(R, W)$ is continuous (and, in fact, convex) for rates greater than the critical rate [9]. In Appendix C, we provide a simple and complete proof of the continuity of the reliability function.

[^0]Lemma 3. Assume that the sequence $\left(R_{n}\right)_{n \in \mathbb{N}}$ converges, and that $R=\lim _{n \rightarrow \infty} R_{n}$. Then we have

$$
\begin{equation*}
\limsup _{n \rightarrow \infty} E\left(n, R_{n}, p\right)=E(R, p) \tag{13}
\end{equation*}
$$

Proof: See Appendix D
Lemma 3 will be pivotal in establishing the exact beeidentification exponent (via Theorem 1), as well as in characterizing the "capacity" of the bee-identification problem (via Theorem 2).

We will characterize the exact bee-identification error exponent for the following scenario:

- For a given $R>0$, the number of bee barcodes $m$ scale exponentially with blocklength $n$ as $m=2^{n R}$.
- For a given $0<\alpha<1$, the number of absentee bees $k$ scale as $k=\lfloor\alpha m\rfloor\rfloor^{3}$ where $\alpha$ denotes the fraction of bees missing from the camera image.
For this scenario, we define the bee-identification exponent as follows ${ }^{4}$

$$
\begin{equation*}
E_{\underline{D}}(R, p, \alpha) \triangleq \limsup _{n \rightarrow \infty}-\frac{1}{n} \log \underline{D}(n, m, p, k) . \tag{14}
\end{equation*}
$$

The following theorem uses Lemmas 1, 2, and 3, to establish the main result in this paper.

Theorem 1. For $0<\alpha<1$, we have

$$
\begin{equation*}
E_{\underline{D}}(R, p, \alpha)=|E(R, p)-R|^{+} \tag{15}
\end{equation*}
$$

where $|x|^{+} \triangleq \max (0, x)$. Further, this exponent is achieved via independent decoding of each barcode.

## Proof: See Appendix E

The above theorem implies the following remarks.
Remark 1. For a given $0<\alpha<1$, if the number of absentee bees $k$ scales as $\alpha m$, then independent barcode decoding is optimal, i.e., independent decoding of barcodes does not lead to any loss in the bee-identification exponent, relative to joint ML decoding of barcodes. This is in contrast to the result in [1], which showed that if no bees are absent, then joint barcode decoding provides significantly better bee-identification exponent relative to independent barcode decoding.
Remark 2. The lower bound on the bee-identification error probability using joint ML decoding in Lemma 2 was obtained by considering only those events in which just a single bee is incorrectly identified. The proof of Theorem 11 employs Lemma 2] and implies that these error events dominate the error exponent.

Remark 3. The bee-identification exponent $E_{\underline{D}}(R, p, \alpha)$ does not depend on the precise value of $0<\alpha<1$.

## A. 'Capacity' of the bee-identification problem

The bee-identification exponent (14) is exactly characterized in terms of the reliability function $E(R, p)$ via Theorem 1 . when the total number of bees scale as $m=2^{n R}$ with $R>$

[^1]0 , and the number of absentee bees scale as $k=\alpha m$ with $0<\alpha<1$. For this same setting, we now formulate and characterize the "capacity" of the bee-identification problem.

For $0 \leq \epsilon<1$, we say that rate $R$ is $(\alpha, \epsilon)$-achievable if $\liminf _{n \rightarrow \infty} \underline{D}\left(n, 2^{n R}, p, \alpha 2^{n R}\right) \leq \epsilon$, and define the $\epsilon$ capacity of the bee-identification problem as the supremum of all $(\alpha, \epsilon)$-achievable rates. We denote this $\epsilon$-capacity as ${ }^{5}$

$$
\begin{equation*}
C_{\underline{D}}(p, \alpha, \epsilon) \triangleq \sup \left\{R: \liminf _{n \rightarrow \infty} \underline{D}\left(n, 2^{n R}, p, \alpha 2^{n R}\right) \leq \epsilon\right\} \tag{16}
\end{equation*}
$$

The above definition implies that for $R<C_{\underline{D}}(p, \alpha, \epsilon)$, there exists a decoding function $\phi$, and a codebook $\overline{\mathcal{C}}$ with $2^{n R}$ codewords having blocklength $n$, for which the bee-identification error probability $D\left(\mathcal{C}, p, \alpha 2^{n R}, \phi\right)<\epsilon$, for infinitely many $n$.

Now, the Bhattacharyya parameter for $\operatorname{BSC}(p)$ is [16]

$$
\begin{equation*}
B_{p} \triangleq-\log \sqrt{4 p(1-p)} \tag{17}
\end{equation*}
$$

and it is well known that [16]

$$
\begin{equation*}
\lim _{R \downarrow 0} E(R, p)=\frac{B_{p}}{2} \tag{18}
\end{equation*}
$$

For a given $0<p<0.5$, define the function $f(R) \triangleq E(R, p)-R$. From (17) and (18), it follows that $\lim _{R \downarrow 0} f(R)>0$, while $f(1)=-1$ because $E(R, p)=0$ for $R \geq 1-H(p)$, where $H(\cdot)$ denotes the binary entropy function. Further, $f(\cdot)$ is continuous because $E(R, p)$ is continuous in $R$ (see Appendix C). Therefore, it follows from the intermediate value theorem [17] that the equation $f(R)=E(R, p)-R=0$ has a positive solution, and this solution is unique because $f(R)$ is strictly decreasing in $R$. The following theorem states that the capacity of the bee-identification problem with absentee bees is equal to the unique solution of the equation $f(R)=0$.
Theorem 2. For $0<\alpha<1$, and every $0 \leq \epsilon<1$, we have

$$
\begin{equation*}
C_{\underline{D}}(p, \alpha, \epsilon)=R_{p}^{*} \tag{19}
\end{equation*}
$$

where $R_{p}^{*}$ is unique positive solution of the following equation

$$
\begin{equation*}
E(R, p)=R \tag{20}
\end{equation*}
$$

Proof: See Appendix F
Theorem 2 and its proof lead to the following remarks.
Remark 4. We prove the strong converse property $[18]$ in Appendix $F$ showing that if $R>R_{p}^{*}$, then the error probability $\underline{D}\left(n, 2^{n R}, p, \alpha 2^{n R}\right)$ tends to 1 as $n \rightarrow \infty$.
Remark 5. The expression for the $\epsilon$-capacity $\sqrt{19}$ is independent of the value of $\alpha \in(0,1)$. Note that a similar behavior was observed for the bee-identification exponent (15).
B. Computation of the bee-identification error exponent $E_{\underline{D}}(R, p, \alpha)$, and the bee-identification capacity $C_{\underline{D}}(p, \alpha, \epsilon)$

We have characterized the bee-identification error exponent $E_{\underline{D}}(R, p, \alpha)$, and capacity $C_{\underline{D}}(p, \alpha, \epsilon)$, via Theorem 1 and Theorem 2, respectively. In this subsection, we discuss some computational aspects of $E_{\underline{D}}(R, p, \alpha)$ and $C_{\underline{D}}(p, \alpha, \epsilon)$.

[^2]It is clear from (15) and 19) that explicit calculations of $E_{\underline{D}}(R, p, \alpha)$ and $C_{\underline{D}}(p, \alpha, \epsilon)$ require the knowledge of the reliability function $\bar{E}(R, p)$ for a range of values of $R$ and $p$. Although the exact value of $E(R, p)$ is not known for all $R$ and $p$, it can be bounded as follows

$$
\begin{equation*}
E_{\mathrm{TLC}}(R, p) \leq E(R, p) \leq E_{\mathrm{sp}}(R, p) \tag{21}
\end{equation*}
$$

where $E_{\mathrm{TLC}}(R, p)$ is the exponent using typical linear codes [19] that achieves the best known lower bound on $E(R, p)$ at all rates, and $E_{\mathrm{sp}}(R, p)$ is the sphere packing exponent [9] for $\operatorname{BSC}(p)$. The exponent $E_{\mathrm{TLC}}(R, p)$ can be explicitly evaluated using the following expression [19]

$$
E_{\mathrm{TLC}}(R, p) \triangleq \begin{cases}\delta_{\mathrm{GV}}(R) B_{p} & 0<R \leq R_{\mathrm{ex}}  \tag{22}\\ R_{0}-R & R_{\mathrm{ex}} \leq R \leq R_{\mathrm{cr}} \\ D\left(\delta_{\mathrm{GV}}(R) \| p\right) & R_{\mathrm{cr}} \leq R \leq 1-H(p)\end{cases}
$$

where $\delta_{\mathrm{GV}}(R)$ is the Gilbert-Varshamov (GV) distance [19] defined as the value of $\delta$ in the interval $[0,0.5]$ with $H(\delta)=$ $1-R$, and

$$
\begin{gathered}
R_{\mathrm{ex}} \triangleq 1-H\left(\frac{\sqrt{4 p(1-p)}}{1+\sqrt{4 p(1-p)}}\right), \\
R_{\mathrm{cr}} \triangleq 1-H\left(\frac{\sqrt{p}}{\sqrt{p}+\sqrt{1-p}}\right) \\
R_{0} \triangleq 1-\log (1+\sqrt{4 p(1-p)}) \\
D(x \| y) \triangleq x \log \frac{x}{y}+(1-x) \log \frac{1-x}{1-y} .
\end{gathered}
$$

The sphere packing exponent is defined as [19]

$$
\begin{equation*}
E_{\mathrm{sp}}(R, p) \triangleq D\left(\delta_{\mathrm{GV}}(R) \| p\right), \quad 0<R \leq 1-H(p) \tag{23}
\end{equation*}
$$

From 21, 22, and 23, we observe that $E(R, p)=$ $D\left(\delta_{\mathrm{GV}}(R) \| p\right)$ for $R_{\mathrm{cr}}<R \leq 1-H(p)$, and it is well known that $E(R, p)$ is identically zero for $R \geq 1-H(p)$ [9]. The exponent $E_{\mathrm{TLC}}(R, p)$ is equal to the random coding exponent $E_{\mathrm{r}}(R, p)$ [9] for $R \geq R_{\mathrm{ex}}$, and therefore the random coding exponent is a tight lower bound on $E(R, p)$ for $R \geq R_{\text {cr }}$. Although it is not so well known, it is also true that for $0.046<p<0.5$, the lower bound $E_{\mathrm{r}}(R, p)$ is tight for certain rates strictly less than the critical rate $R_{\text {cr }}$ [20, Thm. 17].

In general, upper and lower bounds on $E_{\underline{D}}(R, p, \alpha)$ and $C_{\underline{D}}(p, \alpha)$ can be obtained via Theorem 1 and Theorem 2 , respectively, and employing best known bounds on $E(R, p)$.

If we define the following minimum distance metrics

$$
\begin{aligned}
d^{*}(n, R) & \triangleq \max _{C \in \mathscr{C}(n, R)} \min _{\substack{\boldsymbol{c}_{i}, \boldsymbol{c}_{j} \in C \\
c_{i} \neq \boldsymbol{c}_{j}}} \mathrm{~d}_{\mathrm{H}}\left(\boldsymbol{c}_{i}, \boldsymbol{c}_{j}\right) \\
\delta^{*}(n, R) & \triangleq \frac{d^{*}(n, R)}{n} \\
\delta^{*}(R) & \triangleq \limsup _{n \rightarrow \infty} \delta^{*}(n, R)
\end{aligned}
$$

then $E(R, p)$ can also be upper bounded as [16]

$$
\begin{equation*}
E(R, p) \leq \delta^{*}(R) B_{p} \tag{24}
\end{equation*}
$$

The exact value of $\delta^{*}(R)$ is not known in general, though we know that $\delta^{*}(0)=0.5$ and $\delta^{*}(1)=0$ [16]. The value $\delta^{*}(R)$
is lower bounded by $\delta_{\mathrm{GV}}(R)$, and can be upper bounded as follows [21], [22]

$$
\begin{equation*}
\delta^{*}(R) \leq \delta_{\mathrm{LP}}(R) \triangleq \frac{1}{2}-\sqrt{\delta_{\mathrm{GV}}(1-R)\left(1-\delta_{\mathrm{GV}}(1-R)\right)} \tag{25}
\end{equation*}
$$

Combining 24 and 25, we observe that $E(R, p)$ can be upper bounded as follows

$$
\begin{equation*}
E(R, p) \leq \delta_{\mathrm{LP}}(R) B_{p} \tag{26}
\end{equation*}
$$

The following proposition provides an exact and explicit characterization of $E_{\underline{D}}(R, p, \alpha)$ for certain values of $R$ by applying different bounds on $E(R, p)$.

Proposition 1. For given $0<p<0.5$ and $0<\alpha<1$, we have

$$
\begin{equation*}
\lim _{R \downarrow 0} E_{\underline{D}}(R, p, \alpha)=B_{p} / 2 \tag{27}
\end{equation*}
$$

Further, we have $E_{\underline{D}}(R, p, \alpha)=0$ when $R \geq R_{\mathrm{cr}}$.
Proof: From (15), 21), and 22, we obtain

$$
\begin{equation*}
\lim _{R \downarrow 0} E_{\underline{D}}(R, p, \alpha) \geq \lim _{R \downarrow 0} \delta_{\mathrm{GV}}(R) B_{p}=B_{p} / 2 \tag{28}
\end{equation*}
$$

On the other hand, using (15) and (26), we get

$$
\begin{equation*}
\lim _{R \downarrow 0} E_{\underline{D}}(R, p, \alpha) \leq \lim _{R \downarrow 0} \delta_{\mathrm{LP}}(R) B_{p}=B_{p} / 2 \tag{29}
\end{equation*}
$$

and the claim in 27) follows by combining (28) and 29).
Next, we note that $E_{\text {TLC }}\left(R_{\text {cr }}, p\right)=R_{0}-R_{\text {cr }}=$ $D\left(\delta_{\mathrm{GV}}\left(R_{\mathrm{cr}}\right) \| p\right)=E_{\mathrm{sp}}\left(R_{\mathrm{cr}}, p\right)$. Therefore, we have $E\left(R_{\mathrm{cr}}, p\right)=R_{0}-R_{\mathrm{cr}}$, and it follows from 15) that

$$
E_{\underline{D}}\left(R_{\mathrm{cr}}, p, \alpha\right)=\left|R_{0}-2 R_{\mathrm{cr}}\right|^{+}=0
$$

where the last equality follows because $R_{0} \leq 2 R_{\text {cr }}$ [1]. Finally, the fact that $E_{\underline{D}}(R, p, \alpha)$ is non-increasing in $R$ implies that $E_{\underline{D}}(R, p, \alpha)=\overline{0}$ for $R \geq R_{\text {cr }}$. We remark that this result, together with Theorem 2 , implies that $C_{\underline{D}}(p, \alpha)<R_{\text {cr }}$.

It is known that for small rates, the explicit upper bound on $E(R, p)$ given by 26 is better than the sphere packing bound $E_{\mathrm{sp}}(R, p)$ [16]. Further improved upper bounds on $E(R, p)$ can be obtained by using the straight line bound [23], which for $\operatorname{BSC}(p)$ implies that for $R_{1}<R_{2}$, the straight line joining $\delta_{\mathrm{LP}}\left(R_{1}\right) B_{p}$ and $E_{\mathrm{sp}}\left(R_{2}, p\right)$ is an upper bound on $E(R, p)$ for $R \in\left(R_{1}, R_{2}\right)$ [16].

Next, we apply the previously discussed bounds on $E(R, p)$ to compute and explicitly bound $E_{\underline{D}}(R, p, \alpha)$ and $C_{\underline{D}}(p, \alpha)$. Fig. 3 plots upper and lower bounds on $E_{\underline{D}}(R, p, \alpha)$ for $p=$ 0.05 . It is seen from Fig. 3 that the upper and lower bounds coincide as $R$ tends to 0 , and as shown in Prop. 11, we have $\lim _{R \downarrow 0} E_{\underline{D}}(R, p, \alpha)=B_{p} / 2=0.599$ for $p=0.05$.

Fig. 4 plots lower and upper bounds on the bee-identification capacity $C_{\underline{D}}(p, \alpha, \epsilon)$. As shown in Theorem 2, when $0<\alpha<$ 1, the capacity is independent of the value of $\alpha$ (and of $\epsilon$ ). The numerical results in Fig. 4 are obtained using Theorem 2 , and applying the bounds on the reliability function $E(R, p)$, presented in this subsection. It is observed that lower and upper bounds on $C_{\underline{D}}(p, \alpha, \epsilon)$ are relatively close to each other for $p>0.05$.


Fig. 3: Bounds on $E_{\underline{D}}(R, p, \alpha)$ for $p=0.05$.


Fig. 4: Lower and upper bounds on $C_{\underline{D}}(p, \alpha, \epsilon)$.

## C. Curious case of $\lim _{\alpha \downarrow 0} E_{\underline{D}}(R, p, \alpha)$

In this section, we analyze the limiting behavior of $E_{\underline{D}}(R, p, \alpha)$ in the setting where $\alpha \downarrow 0$. We will let $E_{\underline{D}}(R, p)$ be the exponent for the no absentee bee scenario (with $\bar{k}=0$ ), and will compare $E_{\underline{D}}(R, p)$ to $\lim _{\alpha \downarrow 0} E_{\underline{D}}(R, p, \alpha)$.

Now, the exponent $E_{\underline{D}}(R, p)$ was studied in detail in [1] where upper and lower bounds for the same were derived ${ }^{6}$ and it was shown that for $0<R<0.5 R_{\mathrm{ex}}$, we have [1]

$$
\begin{equation*}
2 \delta_{\mathrm{GV}}(2 R) B_{p} \leq E_{\underline{D}}(R, p) \leq 2 \delta_{\mathrm{LP}}(R) B_{p}+R \tag{30}
\end{equation*}
$$

The next theorem shows that $\lim _{\alpha \downarrow 0} E_{\underline{D}}(R, p, \alpha)$ is strictly less than $E_{\underline{D}}(R, p, \alpha)$ at low rates.

Theorem 3. For $0<R<\min \left\{0.169, R_{\mathrm{ex}} / 2\right\}$, we have the following strict inequality

$$
\begin{equation*}
\lim _{\alpha \downarrow 0} E_{\underline{D}}(R, p, \alpha)<E_{\underline{D}}(R, p) . \tag{31}
\end{equation*}
$$

${ }^{6}$ The exponent $E_{\underline{D}}(R, p)$ was defined in 1 Eq. (5)] for $m=2^{n R}$ as $\lim \inf _{n \rightarrow \infty} \frac{1}{n} \log (1 / \underline{D}(n, R, p))$. However, the bounds on $E_{\underline{D}}(R, p)$ presented in [1] continue to hold if liminf is replaced by limsup in the definition of $\underline{E}_{\underline{D}}(R, p)$.

Proof: See Appendix G
The above result highlights that the limiting behavior for the absentee bee scenario, with $\alpha \downarrow 0$, is quite distinct from the scenario where all bees are present. Independent decoding of bee barcodes is optimal for the absentee bee scenario, even when arbitrarily small fraction of bees are absent. On the other hand, for the scenario where all bees are present, strictly better error exponent, than that obtained by independent decoding, can be achieved via joint ML decoding of barcodes [1].

## D. Extension of results to discrete memoryless channels

In the preceding discussion, we characterized the error exponent and capacity for the bee-identification problem with absentee bees under $\operatorname{BSC}(p)$ noise. This characterization can be readily extended to more general discrete memoryless channels (DMCs).

Consider a DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and channel transition matrix $W$. Then, the reliability function of the DMC, denoted $E(R, W)$, is defined as $E(R, W) \triangleq \limsup _{n \rightarrow \infty}-\frac{1}{n} \log P_{\mathrm{e}}\left(n, 2^{n R}, W\right)$, where $P_{\mathrm{e}}\left(n, 2^{n R}, W\right)$ denotes the minimum error probability over all length- $n$ block codes with $2^{n R}$ codewords. Analogous to (4), we may let $\underline{D}(n, m, W, k)$ be the minimum expected beeidentification error probability over the DMC $W$, and define the corresponding bee-identification exponent as

$$
E_{\underline{D}}(R, W, \alpha) \triangleq \limsup _{n \rightarrow \infty}-\frac{1}{n} \log \underline{D}\left(n, 2^{n R}, W, \alpha 2^{n R}\right) .
$$

Let us restrict our attention to DMCs with the property that there exists an output symbol $y \in \mathcal{Y}$ that is reachable from all symbols in $\mathcal{X}$. This property ensures that $E(R, W)$ is continuous for each $R>0$ (see Appendix C). The continuity of $E(R, W)$ can be applied to obtain the following result, equivalent to Theorem 1 for $\operatorname{BSC}(p)$, as

$$
E_{\underline{D}}(R, W, \alpha)=|E(R, W)-R|^{+} .
$$

If, analogous to (16), we define the bee-identification $\epsilon$ capacity over DMC $W$ as

$$
C_{\underline{D}}(W, \alpha, \epsilon) \triangleq \sup \left\{R: \liminf _{n \rightarrow \infty} \underline{D}\left(n, 2^{n R}, W, \alpha 2^{n R}\right) \leq \epsilon\right\}
$$

then, for $0 \leq \epsilon<1$, the $\epsilon$-capacity $C_{\underline{D}}(W, \alpha, \epsilon)$ is equal to the unique positive solution of the equation $E(R, W)=R$. Note that this result also uses the continuity of $E(R, W)$, and extends the result in Theorem 2 to DMCs.

## IV. Reflections

This work extended the characterization of the beeidentification error exponent to the "absentee bees" scenario, where a fraction of the bees are absent in the beehive image. For this scenario, we presented the exact characterization of the bee-identification error exponent in terms of the well known reliability function [9].

The derivation of the bee-identification exponent led to three interesting observations. The first observation is that when the number of absentee bees $k$ scales as $k=\alpha m$, where $\alpha$ lies in the interval $(0,1)$ and is fixed, and the number of bees $m$ scales exponentially with blocklength, then independent
barcode decoding is optimal, i.e., joint decoding of the bee barcodes does not result in any better error exponent relative to the independent decoding of each noisy barcode. This result is in contrast to the result without absentee bees [1], where joint barcode decoding results in significantly higher error exponent compared to independent barcode decoding. The second interesting observation is that when $k=\alpha m$, the beeidentification exponent is dominated by the events where a single bee in the beehive image is incorrectly identified as one of the absentee bees, while the other bee barcodes are correctly decoded. The third observation is that for $k=\alpha m$, the bee-identification exponent does not depend on the actual value of $\alpha$ when $0<\alpha<1$.

We also characterized the exact "capacity" for the beeidentification problem with absentee bees, and proved the strong converse. Further, we showed that for low rates, the error exponent for the case where $\alpha \downarrow 0$ is strictly lower than the corresponding error exponent for the case without absentee bees, thereby highlighting a discontinuity in the error exponent function at $\alpha=0$.

The extension of the results presented in this work to general DMCs was briefly discussed in Section III-D. Future work includes exploring the error exponent for the scenario where $\alpha$, the fraction of absentee bees, also varies with blocklength $n$, and second-order or finite-length analysis, i.e., the scaling of the code rate when $0 \leq \epsilon<1$ and $n$ is finite.

## APPENDIX A

## Proof of Lemma 1

Proof: For a given codebook $\mathcal{C}$, and given $\pi_{(m-k)}$ and $1 \leq i \leq m-k$, the probability $\operatorname{Pr}\left\{\nu(i) \neq \pi_{(m-k)}(i)\right\}$ in (5) is the probability that the codeword $\boldsymbol{c}_{\pi(i)}$ transmitted over $\mathrm{BSC}(p)$ is incorrectly decoded at the receiver. As $\pi_{(m-k)}$ is uniformly distributed over $\Upsilon$, we have for $1 \leq i \leq m-k$,

$$
\begin{equation*}
\min _{\mathcal{C} \in \mathscr{C}(n, m)} \mathbb{E}_{\pi_{(m-k)}}\left[\operatorname{Pr}\left\{\nu(i) \neq \pi_{(m-k)}(i)\right\}\right]=P_{\mathrm{e}}(n, m, p) \tag{32}
\end{equation*}
$$

Now, from the definition of $\underline{D}(n, m, p, k)$ in (4), we get

$$
\begin{align*}
\underline{D}(n, m, p, k) & \leq \min _{\mathcal{C} \in \mathscr{C}(n, m)} D\left(\mathcal{C}, p, k, \phi_{\mathrm{I}}\right) \\
& \leq(m-k) P_{\mathrm{e}}(n, m, p), \tag{33}
\end{align*}
$$

where (33) follows by combining (5) and (32). Finally, the lemma is proved by using (33), and noting that the beeidentification error probability $\underline{D}(n, m, p, k)$ is trivially upper bounded by 1 .

## Appendix B

Proof of Lemma 2
Proof: Let $\mathcal{I}$ denote the image of $\pi_{(m-k)}$, i.e.,

$$
\begin{equation*}
\mathcal{I} \triangleq\left\{j: j=\pi_{(m-k)}(i), 1 \leq i \leq m-k\right\} \tag{34}
\end{equation*}
$$

Let the complement of $\mathcal{I}$ be denoted $\overline{\mathcal{I}}$, i.e.,

$$
\begin{equation*}
\overline{\mathcal{I}} \triangleq\{1,2, \ldots, m\} \backslash \mathcal{I} \tag{35}
\end{equation*}
$$

For a given $\ell \in\{1, \ldots, m-k\}$ and $\jmath \in \overline{\mathcal{I}}$, define the map $\pi_{(m-k)}^{(\ell \rightarrow j)}:\{1, \ldots, m-k\} \rightarrow\{1, \ldots, m\}$ as follows

$$
\pi_{(m-k)}^{(\ell \rightarrow \jmath)}(i) \triangleq \begin{cases}\jmath, & i=\ell \\ \pi_{(m-k)}(i), & i \neq \ell\end{cases}
$$

Thus, the map $\pi_{(m-k)}^{(\ell \rightarrow J)}$ differs with $\pi_{(m-k)}$ only at $i=\ell$, and we have $\pi_{(m-k)}^{(\ell \rightarrow j)} \in \Upsilon$. Now, define the event
$\left\{\pi_{(m-k)} \rightarrow \pi_{(m-k)}^{(\ell \rightarrow \jmath)}\right\} \triangleq\left\{\mathrm{d}_{\mathrm{H}}\left(\tilde{\boldsymbol{c}}_{\ell}, \boldsymbol{c}_{\jmath}\right) \leq \mathrm{d}_{\mathrm{H}}\left(\tilde{\boldsymbol{c}}_{\ell}, \boldsymbol{c}_{\pi_{(m-k)}(\ell)}\right)\right\}$,
where the error event $\left\{\pi_{(m-k)} \rightarrow \pi_{(m-k)}^{(\ell \rightarrow j)}\right\}$ implies that only a single bee is decoded incorrectly. Further, if we define the event $\mathscr{E}_{\pi_{(m-k)}}^{(\ell)}$ as

$$
\begin{equation*}
\mathscr{E}_{\pi_{(m-k)}}^{(\ell)} \triangleq \bigcup_{\jmath \in \overline{\mathcal{I}}}\left\{\pi_{(m-k)} \rightarrow \pi_{(m-k)}^{(\ell \rightarrow \jmath)}\right\} \tag{36}
\end{equation*}
$$

then, for $\ell \in\{1, \ldots, m-k\}$, we have

$$
\begin{align*}
\operatorname{Pr} & \left\{\mathscr{E}_{\pi_{(m-k)}^{(\ell)}}^{(\ell)}\right\} \\
& =\operatorname{Pr}\left\{\bigcup_{\jmath \in \overline{\mathcal{I}}}\left\{\mathrm{d}_{\mathrm{H}}\left(\tilde{\boldsymbol{c}}_{\ell}, \boldsymbol{c}_{\jmath}\right) \leq \mathrm{d}_{\mathrm{H}}\left(\tilde{\boldsymbol{c}}_{\ell}, \boldsymbol{c}_{\pi_{(m-k)}(\ell)}\right)\right\}\right\} \tag{37}
\end{align*}
$$

The output of the ML decoding function $\phi_{\mathrm{J}}$ is given by (7), and hence the bee-identification error probability is lower bounded as follows

$$
\begin{equation*}
\operatorname{Pr}\left\{\nu \neq \pi_{(m-k)}\right\} \geq \operatorname{Pr}\left\{\bigcup_{1 \leq \ell \leq m-k} \mathscr{E}_{\pi_{(m-k)}}^{(\ell)}\right\} \tag{38}
\end{equation*}
$$

For a given codebook $\mathcal{C}$, we observe from (37) that the event $\mathscr{E}_{\pi_{(m-k)}(\ell)}$ depends only on the noise in the $\ell$-th received barcode $\tilde{\boldsymbol{c}}_{\ell}$. Thus, the set of events $\mathscr{E}_{\pi_{(m-k)}(\ell)}^{(\ell)}$, for $\ell \in\{1, \ldots, m-k\}$, are mutually independent. Therefore, the probability of their union can be lower bounded using Shulman's lower bound as [24, Eq. (30)], [25, p. 109]

$$
\begin{equation*}
\operatorname{Pr}\left\{\bigcup_{1 \leq \ell \leq m-k} \mathscr{E}_{\pi_{(m-k)}(\ell)}\right\} \geq \frac{1}{2} \cdot \min \left\{1, \sum_{\ell=1}^{m-k} \operatorname{Pr}\left\{\mathscr{E}_{\pi_{(m-k)}^{(\ell)}}\right\}\right\} \tag{39}
\end{equation*}
$$

As $|\overline{\mathcal{I}}|=k$, we observe from 37) that the error event $\mathscr{E}_{\pi_{(m-k)}(\ell)}^{(\ell)}$ occurs when the received word $\tilde{\boldsymbol{c}}_{\ell}$ is incorrectly decoded to one of the $k$ incorrect codewords $\left\{\boldsymbol{c}_{\jmath}\right\}_{\jmath \in \overline{\mathcal{I}}}$, instead of the correct codeword $\boldsymbol{c}_{\pi_{(m-k)}(\ell)}$, and so

$$
\begin{equation*}
\mathbb{E}_{\pi_{(m-k)}}\left[\operatorname{Pr}\left\{\mathscr{E}_{\pi_{(m-k)}^{(\ell)}}\right\}\right] \geq P_{\mathrm{e}}(n, k+1, p) \geq P_{\mathrm{e}}(n, k, p), \tag{40}
\end{equation*}
$$

because $P_{\mathrm{e}}(n, k+1, p)$ denotes the average error probability, minimized over all codebooks with only $k+1$ codewords. Note that 40 holds for all codebooks $\mathcal{C} \in \mathscr{C}(n, m)$. Now recall that $\Upsilon$ is the set of all injective maps from $\{1, \ldots, m-k\}$ to $\{1, \ldots, m\}$, and that $\pi_{(m-k)}$ is uniformly distributed over $\Upsilon$. Let $0<\varepsilon<1 / 2$ and $k>1 / \varepsilon$, and define the set

$$
\mathcal{A}^{(\ell)} \triangleq\left\{\pi_{(m-k)} \in \Upsilon: \operatorname{Pr}\left\{\mathscr{E}_{\pi_{(m-k)}^{(\ell)}}^{(\ell)}\right\} \geq P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)\right\}
$$

A key observation is that the size $\left|\mathcal{A}^{(\ell)}\right|$ can be bounded as

$$
\begin{equation*}
\left|\mathcal{A}^{(\ell)}\right|>(1-\varepsilon)|\Upsilon|, \tag{41}
\end{equation*}
$$

where the inequality holds for all $\ell \in\{1, \ldots, m-k\}$, and all codebooks $\mathcal{C} \in \mathscr{C}(n, m)$. This claim can be explained as follows. First, fix the following variables: codebook $\mathcal{C} \in$ $\mathscr{C}(n, m)$, index $\ell \in\{1, \ldots, m-k\}$, and $\pi_{(m-k)} \in \Upsilon$. Note that fixing $\pi_{(m-k)}$ in turn fixes the sets $\mathcal{I}$ and $\overline{\mathcal{I}}$, defined in (34) and (35), respectively. Let $\overline{\mathcal{I}}=\left\{\jmath_{1}, \ldots, \jmath_{k}\right\}$, and define

$$
q_{\jmath_{r}}^{(\ell)}=\operatorname{Pr}\left\{\mathscr{E}_{\substack{\left(\ell \rightarrow \jmath^{\prime} r\right) \\ \pi_{(m-k)}^{(\ell)}}}^{(\ell)} \text { for } r \in\{1, \ldots, k\}\right.
$$

Now, let $\left(\tilde{\jmath}_{1} \tilde{\jmath}_{2} \ldots \tilde{\jmath}_{k}\right)$ be a permutation of the indices $\left(\jmath_{1} \jmath_{2} \ldots \jmath_{k}\right)$ such that we have the following relation

$$
q_{\tilde{\jmath}_{1}}^{(\ell)} \leq q_{\tilde{\jmath}_{2}}^{(\ell)} \leq \cdots \leq q_{\tilde{\jmath}_{k}}^{(\ell)} .
$$

Thus, $q_{\tilde{J}_{\lfloor k \varepsilon\rfloor}^{(\ell)}}^{(\ell)}$ satisfies the following property

$$
q_{\tilde{\jmath}_{\lfloor k \varepsilon\rfloor}}^{(\ell)}=\max \left\{q_{\tilde{\jmath}_{1}}^{(\ell)}, q_{\tilde{\jmath}_{2}}^{(\ell)}, \ldots, q_{\tilde{\jmath}_{\lfloor k \varepsilon\rfloor}}^{(\ell)}\right\}
$$

and hence

$$
q_{\tilde{\jmath}\lfloor k \varepsilon\rfloor}^{(\ell)} \geq P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)
$$

The above relation is satisfied because each of $q_{\tilde{\jmath}_{1}}^{(\ell)}, q_{\tilde{\jmath}_{2}}^{(\ell)}, \ldots, q_{\tilde{\jmath}_{\lfloor k \varepsilon\rfloor}}^{(\ell)}$ is obtained by comparison among codewords, $\boldsymbol{c}_{\pi_{(m-k)}(\ell)}, \boldsymbol{c}_{\jmath_{1}}, \boldsymbol{c}_{\jmath_{2}}, \ldots, \boldsymbol{c}_{\jmath_{k}}$, while $P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)$ is the minimum achievable average error probability using only $\lfloor k \varepsilon\rfloor$ codewords. Therefore, the fraction of entries in the set $\left\{q_{\jmath_{1}}^{(\ell)}, q_{\jmath_{2}}^{(\ell)}, \ldots, q_{\jmath_{k}}^{(\ell)}\right\}$ that satisfy

$$
\left.q_{J_{r}}^{(\ell)}=\operatorname{Pr}\left\{\mathscr{E}_{\mathscr{E}_{(m-k)}^{(\ell)}}^{\pi_{\left(m-J_{r}\right)}^{(\ell)}}\right\}\right\} \geq P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p),
$$

is at least $(k-\lfloor k \varepsilon\rfloor+1) / k>(1-\varepsilon)$. This technique can be reapplied to other mappings in $\Upsilon$ to obtain 41.

Next, construct a matrix $\mathcal{M}$ whose rows are indexed by elements of $\Upsilon$, and the columns are indexed by $\ell \in\{1,2, \ldots, m-$ $k\}$. For a given row of $\mathcal{M}$ indexed by $\pi_{(m-k)} \in \Upsilon$, let the $\ell$-th entry be equal to $\operatorname{Pr}\left\{\mathscr{E}_{\pi_{(m-k)}(\ell)}\right\}$. Then, from (41) it follows that at least $1-\varepsilon$ fraction of entries in each column are lower bounded by $P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)$. Thus, the fraction of entries of matrix $\mathcal{M}$ that are lower bounded by $P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)$ is at least $1-\varepsilon$. Now, we call a row of $\mathcal{M}$ to be $\varepsilon$-strong if the fraction of entries lower bounded by $P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)$ in that row exceed $\varepsilon$. Let $\theta_{\varepsilon}$ denote the fraction of rows of $\mathcal{M}$ that are $\varepsilon$-strong. Then we have

$$
\left(1-\theta_{\varepsilon}\right) \varepsilon+\theta_{\varepsilon} \geq 1-\varepsilon
$$

which implies that

$$
\begin{equation*}
\theta_{\varepsilon} \geq \frac{1-2 \varepsilon}{1-\varepsilon}>1-2 \varepsilon \tag{42}
\end{equation*}
$$

Now, we define

$$
\begin{align*}
\Upsilon_{\varepsilon} & \triangleq\left\{\pi_{(m-k)} \in \Upsilon:\right. \\
& \left.\sum_{\ell=1}^{m-k} \operatorname{Pr}\left\{\mathscr{E}_{\pi_{(m-k)}}^{(\ell)}\right\}>(m-k) \varepsilon P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)\right\} \tag{43}
\end{align*}
$$

and note that the elements of $\Upsilon_{\varepsilon}$ correspond to the rows of $\mathcal{M}$ whose row-sum is greater than $(m-k) \varepsilon P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)$. Now, we observe that if $\pi_{(m-k)}$ corresponds to a row of $\mathcal{M}$ that is $\varepsilon$-strong, then $\pi_{(m-k)} \in \Upsilon_{\varepsilon}$. Therefore, we have

$$
\begin{equation*}
\left|\Upsilon_{\varepsilon}\right| \geq \theta_{\varepsilon}|\Upsilon|>(1-2 \varepsilon)|\Upsilon| \tag{44}
\end{equation*}
$$

where the strict inequality follows from (42). Finally, we have

$$
\begin{align*}
& \mathbb{E}_{\pi_{(m-k)}}\left[\operatorname{Pr}\left\{\nu \neq \pi_{(m-k)}\right\}\right] \\
& \stackrel{(\mathrm{i})}{=} \frac{1}{|\Upsilon|} \sum_{\pi_{(m-k)} \in \Upsilon} \operatorname{Pr}\left\{\nu \neq \pi_{(m-k)}\right\}, \\
& \stackrel{(\mathrm{ii)}}{\geq} \frac{1}{|\Upsilon|} \sum_{\pi_{(m-k)} \in \Upsilon} \frac{1}{2} \cdot \min \left\{1, \sum_{\ell=1}^{m-k} \operatorname{Pr}\left\{\mathscr{E}_{\pi_{(m-k)}(\ell)}\right\}\right\} \\
& \quad \stackrel{\text { (iii) }}{\geq} \frac{1}{|\Upsilon|} \sum_{\pi_{(m-k)} \in \Upsilon_{\varepsilon}} \frac{1}{2} \cdot \min \left\{1, \sum_{\ell=1}^{m-k} \operatorname{Pr}\left\{\mathscr{E}_{\pi_{(m-k)}(\ell)}\right\}\right\} \\
& \quad \text { (iv) } \frac{1}{\geq} \sum_{\pi_{(m-k)} \in \Upsilon_{\varepsilon}} \frac{1}{2} \cdot \min \left\{1,(m-k) \varepsilon P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)\right\}, \\
& \quad \text { (v) } \frac{1-2 \varepsilon}{>} \cdot \min \left\{1,(m-k) \varepsilon P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)\right\}, \tag{45}
\end{align*}
$$

where (i) follows because $\pi_{(m-k)}$ is uniformly distributed over $\Upsilon$, (ii) follows from combining (38) and (39), (iii) follows because we restrict $\pi_{(m-k)}$ to belong to $\Upsilon_{\varepsilon} \subseteq \Upsilon$, (iv) follows using (43) as $\sum_{\ell=1}^{m-k} \operatorname{Pr}\left\{\mathscr{E}_{(m-k)}^{(\ell)}\right\}>(m-k) \varepsilon P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)$ for every $\pi_{(m-k)} \in \Upsilon_{\varepsilon}$, and (v) follows from the fact that $\left|\Upsilon_{\varepsilon}\right| /|\Upsilon|>1-2 \varepsilon$ via (44). Finally, we obtain (9) by combining (3), (8), with the fact that (45) holds for all codebooks $\mathcal{C} \in \mathscr{C}(n, m)$.

We now proceed to prove the alternative (and stronger) bound in 10. Choose a codebook $\mathcal{C} \in \mathscr{C}(n, m)$ and a mapping $\pi_{(m-k)} \in \Upsilon$. For $\ell \in\{1, \ldots, m-k\}$, let the error event $\mathscr{E}_{\pi_{(m-k)}}^{(\ell)}$ be given by 36. Applying (38), we get

$$
\begin{align*}
& \operatorname{Pr}\left\{\nu \neq \pi_{(m-k)}\right\} \\
& \geq \operatorname{Pr}\left\{\bigcup_{1 \leq \ell \leq m-k} \mathscr{E}_{\pi_{(m-k)}(\ell)}\right\} \\
& \stackrel{(\mathrm{vi})}{=} 1-\prod_{\ell=1}^{m-k}\left(1-\operatorname{Pr}\left\{\mathscr{E}_{\pi_{(m-k)}^{(\ell)}}^{(\ell-1}\right\}\right) \\
& =1-\exp \left(\sum_{\ell=1}^{m-k} \ln \left(1-\operatorname{Pr}\left\{\mathscr{E}_{\pi_{(m-k)}^{(\ell)}}^{(\ell)}\right\}\right)\right) \\
& \stackrel{(\mathrm{vii})}{\geq} 1-\exp \left(-\sum_{\ell=1}^{m-k} \operatorname{Pr}\left\{\mathscr{E}_{\pi_{(m-k)}^{(\ell)}}\right\}\right) \tag{46}
\end{align*}
$$

where (vi) follows because the events $\mathscr{E}_{\pi_{(m-k)}}^{(\ell)}$, for $\ell \in$ $\{1, \ldots, m-k\}$, are mutually independent, and (vii) follows because $\ln (1-x) \leq-x$ for $x \in[0,1)$. Now, we have

$$
\begin{aligned}
& \mathbb{E}_{\pi_{(m-k)}}\left[\operatorname{Pr}\left\{\nu \neq \pi_{(m-k)}\right\}\right] \\
& \quad=\frac{1}{|\Upsilon|} \sum_{\pi_{(m-k)} \in \Upsilon} \operatorname{Pr}\left\{\nu \neq \pi_{(m-k)}\right\}
\end{aligned}
$$

$$
\begin{align*}
& \stackrel{\text { (viii) }}{\geq} \frac{1}{|\Upsilon|} \sum_{\pi_{(m-k)} \in \Upsilon}\left[1-\exp \left(-\sum_{\ell=1}^{m-k} \operatorname{Pr}\left\{\mathscr{E}_{\pi_{(m-k)}(\ell)}^{(\ell)}\right\}\right)\right] \\
& \geq \frac{1}{|\Upsilon|} \sum_{\pi_{(m-k)} \in \Upsilon_{\varepsilon}}\left[1-\exp \left(-\sum_{\ell=1}^{m-k} \operatorname{Pr}\left\{\mathscr{E}_{\pi_{(m-k)}(\ell)}^{(\ell)}\right\}\right)\right] \\
& \stackrel{\text { (ix) }}{\geq} \frac{1}{|\Upsilon|} \sum_{\pi_{(m-k)} \in \Upsilon_{\varepsilon}}\left[1-\exp \left(-(m-k) \varepsilon P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)\right)\right] \\
& >(1-2 \varepsilon)\left[1-\exp \left(-(m-k) \varepsilon P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)\right)\right] \tag{47}
\end{align*}
$$

where (viii) follows from (46, and (ix) follows using (43) as $\sum_{\ell=1}^{m-k} \operatorname{Pr}\left\{\mathscr{E}_{(m-k)}^{(\ell)}\right\}>(m-k) \varepsilon P_{\mathrm{e}}(n,\lfloor k \varepsilon\rfloor, p)$ for every $\pi_{(m-k)} \in \Upsilon_{\varepsilon}$. Finally, combining (3), (8), with the fact that (47) holds for all codebooks $\mathcal{C} \in \mathscr{C}(n, m)$, we obtain the lower bound on $\underline{D}(n, m, p, k)$ in 10 .

We remark that 10 provides a strict improvement over the corresponding bound in (9). This is because the bound in (10) exploits mutual independence of events $\left\{\pi_{(m-k)} \rightarrow \pi_{(m-k)}^{(\ell \rightarrow j)}\right\}$, while the bound in (9) only uses their pairwise independence.

## Appendix C

## Continuity of the Reliability Function

Consider a DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and transition matrix $W$. Then, the (operationally defined) reliability function of the DMC is defined as [9]

$$
\begin{equation*}
E(R, W) \triangleq \limsup _{n \rightarrow \infty}-\frac{1}{n} \log P_{\mathrm{e}}\left(n, 2^{n R}, W\right) \tag{48}
\end{equation*}
$$

where $P_{\mathrm{e}}\left(n, 2^{n R}, W\right)$ denotes the minimum error probability over all length- $n$ block codes with $2^{n R}$ codewords.

We prove that $E(R, W)$ is continuous at any $R>R_{\infty}$, where $R_{\infty}$ is defined as the smallest $R \geq 0$ for which the sphere-packing exponent $E_{\mathrm{sp}}(R, W)$ is finite [26, p. 69]. Note that $R_{\infty}>0$ if and only if each output $y \in \mathcal{Y}$ is unreachable from at least one input symbol in $\mathcal{X}$ [26, p. 70], and hence $R_{\infty}=0$ for $\operatorname{BSC}(p)$. In the following, for brevity, we suppress the dependence of several quantities on $W$, e.g., we denote $E(R, W), E_{\mathrm{sp}}(R, W)$, and $P_{\mathrm{e}}\left(n, 2^{n R}, W\right)$ by $E(R), E_{\mathrm{sp}}(R)$, and $P_{\mathrm{e}}\left(n, 2^{n R}\right)$, respectively.

Lemma 4. The reliability function $E(R)$ is continuous at any $R>R_{\infty}$.

Proof: Let $P_{\mathrm{e}}(n, M, L)$ denote the minimum error probability for the given channel minimized over all codes with $M$ code words of length $n$ and all list decoding schemes with list size $L$. Then, [26, Thm. 1] states that

$$
\begin{equation*}
P_{\mathrm{e}}\left(n_{1}+n_{2}, M, L_{2}\right) \geq P_{\mathrm{e}}\left(n_{1}, M, L_{1}\right) P_{\mathrm{e}}\left(n_{2}, L_{1}+1, L_{2}\right) \tag{49}
\end{equation*}
$$

Note that when the list size is $L=1$, then the list-decoding error corresponds to ordinary decoding error, and for $M=$ $2^{n R}$, we have $P_{\mathrm{e}}(n, M, L=1)=P_{\mathrm{e}}(n, M)=P_{\mathrm{e}}\left(n, 2^{n R}\right)$.

We will employ (49) for proving the continuity of $E$ at a point $R>R_{\infty}$. Fix $\delta \in(0,1)$ and let $n_{1}=\delta n, n_{2}=(1-\delta) n$, $L_{2}=1$ in (49) to obtain

$$
\begin{align*}
P_{\mathrm{e}}(n, M) & \geq P_{\mathrm{e}}\left(\delta n, M, L_{1}\right) P_{\mathrm{e}}\left((1-\delta) n, L_{1}+1\right), \\
& \geq P_{\mathrm{e}}\left(\delta n, M, L_{1}\right) P_{\mathrm{e}}\left((1-\delta) n, L_{1}\right) \tag{50}
\end{align*}
$$

Let $M=2^{n R}, R^{\prime}=R-R_{\infty}$, and $L_{1}=2^{n(1-\delta)\left(R+\delta R^{\prime}\right)}$. Then $M / L_{1}=2^{\delta n\left(R_{\infty}+\delta R^{\prime}\right)}$ and it follows from [26, Thm. 2]

$$
\begin{equation*}
\limsup _{n \rightarrow \infty}-\frac{1}{n} \log P_{\mathrm{e}}\left(\delta n, M, L_{1}\right) \leq \delta E_{\mathrm{sp}}\left(R_{\infty}+\delta R^{\prime}\right) \tag{51}
\end{equation*}
$$

As $L_{1}=2^{(1-\delta) n\left(R+\delta R^{\prime}\right)}$, it follows from (48) that

$$
\begin{equation*}
\limsup _{n \rightarrow \infty}-\frac{1}{n} \log P_{\mathrm{e}}\left((1-\delta) n, L_{1}\right)=(1-\delta) E\left(R+\delta R^{\prime}\right) \tag{52}
\end{equation*}
$$

The limsup operator is subadditive, i.e., for two sequences $\left(a_{n}\right)$ and $\left(b_{n}\right)$, we have $\lim \sup _{n \rightarrow \infty}\left(a_{n}+b_{n}\right) \leq$ $\limsup _{n \rightarrow \infty} a_{n}+\limsup \sin _{n \rightarrow \infty} b_{n}$. Combining this fact with (50), we get

$$
\begin{aligned}
\limsup _{n \rightarrow \infty}-\frac{1}{n} \log P_{\mathrm{e}}(n, M) & \leq \limsup _{n \rightarrow \infty}-\frac{1}{n} \log P_{\mathrm{e}}\left(\delta n, M, L_{1}\right) \\
& +\limsup _{n \rightarrow \infty}-\frac{1}{n} \log P_{\mathrm{e}}\left((1-\delta) n, L_{1}\right)
\end{aligned}
$$

Applying (51) and 52) to the above inequality, we get

$$
\begin{equation*}
E(R) \leq \delta E_{\mathrm{sp}}\left(R_{\infty}+\delta R^{\prime}\right)+(1-\delta) E\left(R+\delta R^{\prime}\right) \tag{53}
\end{equation*}
$$

As (53) holds for all $\delta>0$, we have

$$
\begin{align*}
E(R) & \leq \liminf _{\delta \downarrow 0}\left[\delta E_{\mathrm{sp}}\left(R_{\infty}+\delta R^{\prime}\right)+(1-\delta) E\left(R+\delta R^{\prime}\right)\right] \\
& \stackrel{(\mathrm{i})}{=} \liminf _{\delta \downarrow 0}\left[(1-\delta) E\left(R+\delta R^{\prime}\right)\right] \\
& =\liminf _{\delta \downarrow 0} E\left(R+\delta R^{\prime}\right) \tag{54}
\end{align*}
$$

where (i) follows because $\lim _{\delta \downarrow 0}\left[\delta E_{\mathrm{sp}}\left(R_{\infty}+\delta R^{\prime}\right)\right]=0$ as $E_{\mathrm{sp}}\left(R_{\infty}\right)$ is finite. Next, as $E(R)$ is non-increasing in $R$,

$$
\begin{equation*}
E\left(R+\delta R^{\prime}\right) \leq E(R) \tag{55}
\end{equation*}
$$

and it follows from (55) that

$$
\begin{equation*}
\limsup _{\delta \downarrow 0} E\left(R+\delta R^{\prime}\right) \leq E(R) \tag{56}
\end{equation*}
$$

Now (54) and (56) imply that $\lim _{\delta \downarrow 0} E\left(R+\delta R^{\prime}\right)$ exists, and

$$
\begin{equation*}
E(R)=\lim _{\delta \downarrow 0} E\left(R+\delta R^{\prime}\right)=\lim _{\delta \downarrow 0} E(R+\delta) \tag{57}
\end{equation*}
$$

The above argument leading to (57) shows that $E(R)$ is right continuous. To prove left continuity, we again fix $R>R_{\infty}$ and choose $\delta \in(0,1)$. Let $R^{\prime}=R-R_{\infty}$, $M=2^{n\left(R_{\infty}+\left[R^{\prime} /(1+\delta)\right]\right)}$ and $L_{1}=2^{n(1-\delta)\left(R_{\infty}+R^{\prime}\right)}$. Then, $M / L_{1}=2^{(\delta n)\left(R_{\infty}+R^{\prime} \delta /(1+\delta)\right)}$. From [26. Thm. 2], we have
$\limsup _{n \rightarrow \infty}-\frac{1}{n} \log P_{\mathrm{e}}\left(\delta n, M, L_{1}\right) \leq \delta E_{\mathrm{sp}}\left(R_{\infty}+R^{\prime} \delta /(1+\delta)\right)$.
As $L_{1}=2^{(1-\delta) n R}$, it follows from (48) that

$$
\begin{equation*}
\limsup _{n \rightarrow \infty}-\frac{1}{n} \log P_{\mathrm{e}}\left((1-\delta) n, L_{1}\right)=(1-\delta) E(R) \tag{59}
\end{equation*}
$$

Combining (50), 58), and 59), we obtain
$E\left(R_{\infty}+\frac{R^{\prime}}{1+\delta}\right) \leq \delta E_{\mathrm{sp}}\left(R_{\infty}+\frac{R^{\prime} \delta}{1+\delta}\right)+(1-\delta) E(R)$.
The above inequality can be equivalently expressed as
$E(R) \geq \frac{1}{1-\delta} E\left(R_{\infty}+\frac{R^{\prime}}{1+\delta}\right)-\frac{\delta}{1-\delta} E_{\mathrm{sp}}\left(R_{\infty}+\frac{R^{\prime} \delta}{1+\delta}\right)$.

The above relation holds for all $\delta>0$, and hence

$$
\begin{aligned}
E(R) \geq \limsup _{\delta \downarrow 0} & \left(\frac{1}{1-\delta} E\left(R_{\infty}+\frac{R^{\prime}}{1+\delta}\right)\right. \\
& \left.-\frac{\delta}{1-\delta} E_{\mathrm{sp}}\left(R_{\infty}+\frac{R^{\prime} \delta}{1+\delta}\right)\right)
\end{aligned}
$$

Now, $\lim _{\delta \downarrow 0}(\delta /(1-\delta)) E_{\text {sp }}\left(R_{\infty}+R^{\prime} \delta /(1+\delta)\right)=0$ as $E_{\mathrm{sp}}\left(R_{\infty}\right)$ is finite, and it follows from above inequality that

$$
\begin{align*}
E(R) & \geq \limsup _{\delta \downarrow 0}\left(\frac{1}{1-\delta} E\left(R_{\infty}+\frac{R^{\prime}}{1+\delta}\right)\right) \\
& =\limsup _{\delta \downarrow 0} E\left(R_{\infty}+\frac{R^{\prime}}{1+\delta}\right) \\
& =\underset{\delta \downarrow 0}{\limsup } E\left(R_{\infty}+R^{\prime}(1-\delta)\right) \\
& =\underset{\delta \downarrow 0}{\lim \sup } E(R-\delta) \tag{60}
\end{align*}
$$

Further, as $E(R)$ is non-increasing in $R$, we have

$$
\begin{equation*}
\liminf _{\delta \downarrow 0} E(R-\delta) \geq E(R) \tag{61}
\end{equation*}
$$

Combining 60) and 61, we get

$$
\begin{equation*}
E(R)=\lim _{\delta \downarrow 0} E(R-\delta) \tag{62}
\end{equation*}
$$

which proves that $E(R)$ is left continuous, and the proof is complete by combining (57) and 62).

## Appendix D

## Proof of Lemma 3

Proof: Choose $R>0$ and $\epsilon>0$. As $E(\tilde{R}, p)$ is continuous in $\tilde{R}$ (see Appendix C), there exists $0<\delta<R$ such that $|E(\tilde{R}, p)-E(R, p)|<\epsilon$ for all $|\tilde{R}-R| \leq \delta$. Now, as $R_{n}$ converges to $R$, there exists an $N$ such that $\left|R_{n}-R\right| \leq \delta$ for all $n \geq N$. Furthermore, as $E(n, R, p)$ is non-increasing in $R$,

$$
E\left(n, R_{n}, p\right) \leq E(n, R-\delta, p), \text { for } n \geq N
$$

From the above inequality, it follows that

$$
\begin{align*}
\limsup _{n \rightarrow \infty} E\left(n, R_{n}, p\right) & \leq \limsup _{n \rightarrow \infty} E(n, R-\delta, p) \\
& =E(R-\delta, p) \\
& <E(R, p)+\epsilon \tag{63}
\end{align*}
$$

As $E(n, R, p)$ is non-increasing in $R$, we have

$$
E\left(n, R_{n}, p\right) \geq E(n, R+\delta, p), \text { for } n \geq N
$$

From the above inequality, it follows that

$$
\begin{align*}
\limsup _{n \rightarrow \infty} E\left(n, R_{n}, p\right) & \geq \limsup _{n \rightarrow \infty} E(n, R+\delta, p) \\
& =E(R+\delta, p) \\
& \geq E(R, p)-\epsilon \tag{64}
\end{align*}
$$

The proof is complete by observing that (63) and 64 hold for all $\epsilon>0$.

## Appendix E Proof of Theorem 1

Proof: We first show that $E_{\underline{D}}(R, p, \alpha) \geq|E(R, p)-R|^{+}$. Towards this, we note that when $m=2^{n R}$ and $k=\alpha m$, for a given $0<\alpha<1$, then we have

$$
\begin{align*}
\lim _{n \rightarrow \infty}-\frac{1}{n} \log (m-k) & =\left(\lim _{n \rightarrow \infty}-\frac{1}{n} \log (1-\alpha)\right)-R \\
& =-R \tag{65}
\end{align*}
$$

Combining (6), (14), and 65), we get

$$
\begin{align*}
E_{\underline{D}}(R, p, \alpha) & \geq\left|\left(\limsup _{n \rightarrow \infty}-\frac{1}{n} \log P_{\mathrm{e}}(n, m, p)\right)-R\right|^{+} \\
& =|E(R, p)-R|^{+} \tag{66}
\end{align*}
$$

where the last equality follows from (12) and (13).
Next, we show that $E_{\underline{D}}(R, p, \alpha) \leq|E(R, p)-R|^{+}$by applying Lemma 2 Choose $\varepsilon=1 / 4$, and define

$$
\begin{equation*}
\hat{R}_{n} \triangleq \frac{1}{n} \log (\lfloor k \varepsilon\rfloor) \tag{67}
\end{equation*}
$$

For $k>8$, we have $k \varepsilon / 2<\lfloor k \varepsilon\rfloor \leq k \varepsilon$. Thus, when $k=\alpha m$ and $m=2^{n R}$, we get

$$
\begin{equation*}
R+\frac{1}{n} \log \left(\frac{\alpha \varepsilon}{2}\right)<\hat{R}_{n} \leq R+\frac{1}{n} \log (\alpha \varepsilon) \tag{68}
\end{equation*}
$$

which implies that

$$
\begin{equation*}
R=\lim _{n \rightarrow \infty} \hat{R}_{n} \tag{69}
\end{equation*}
$$

Combining the facts that $\lim _{n \rightarrow \infty} \frac{1}{n} \log ((1-2 \varepsilon) / 2)=0$, $\lim _{n \rightarrow \infty} \frac{1}{n} \log ((m-k) \varepsilon)=R$, with (14), (9), (12), and (67), we get

$$
\begin{align*}
E_{\underline{D}}(R, p, \alpha) & \leq\left|\limsup _{n \rightarrow \infty} E\left(n, \hat{R}_{n}, p\right)-R\right|^{+} \\
& =|E(R, p)-R|^{+} \tag{70}
\end{align*}
$$

where the last equality follows from $\sqrt{69}$ and (13). The proof is now complete by combining (66) and (70).

## Appendix F

## Proof of Theorem 2

Proof: We first prove the direct part $C_{\underline{D}}(p, \alpha, \epsilon) \geq R_{p}^{*}$. If $R<R_{p}^{*}$, then it follows from (15) and the definition of $R_{p}^{*}$ that $E_{\underline{D}}(R, p, \alpha)$ is strictly positive. From (14) it follows that there exist infinitely many $n$ for which

$$
-\frac{1}{n} \log \underline{D}\left(n, 2^{n R}, p, \alpha 2^{n R}\right)>E_{\underline{D}}(R, p, \alpha) / 2 .
$$

In other words, when $R<R_{p}^{*}, \underline{D}\left(n, 2^{n R}, p, \alpha 2^{n R}\right)<$ $2^{-n E_{\underline{D}}(R, p, \alpha) / 2}$. Thus when $R<R_{p}^{*}$, we have

$$
\liminf _{n \rightarrow \infty} \underline{D}\left(n, 2^{n R}, p, \alpha 2^{n R}\right)=0
$$

Therefore, any rate less than $R_{p}^{*}$ is achievable and it follows from the definition of capacity in (16) that $C_{\underline{D}}(p, \alpha, \epsilon) \geq R_{p}^{*}$.

Next, we will apply the bound in 10 to prove the converse part $C_{\underline{D}}(p, \alpha, \epsilon) \leq R_{p}^{*}$. This is a strong converse statement, i.e., for rates $R>R_{p}^{*}$, the error probability $\underline{D}\left(n, 2^{n R}, p, \alpha 2^{n R}\right)$ tends to 1 as $n \rightarrow \infty$. Consider a rate
$R$ that satisfies $R>R_{p}^{*}$, and define $\Delta_{R} \triangleq R-E(R, p)$. Then it follows from the definition of $R_{p}^{*}$, and the fact $E(R, p)$ is non-increasing in $R$, that $\Delta_{R}>0$. Define $\varepsilon_{n} \triangleq \frac{1}{n}$, and let $n$ be sufficiently large such that $k=\alpha 2^{n R}>2 n=2 / \varepsilon_{n}$. Now define $\hat{R}_{n}$ to be

$$
\begin{equation*}
\hat{R}_{n} \triangleq \frac{1}{n} \log \left(\left\lfloor k \varepsilon_{n}\right\rfloor\right) \tag{71}
\end{equation*}
$$

Then, we have

$$
\begin{align*}
R+\frac{1}{n} \log \left(\frac{\alpha}{2 n}\right) & <\hat{R}_{n} \leq R+\frac{1}{n} \log \left(\frac{\alpha}{n}\right) \\
R & =\lim _{n \rightarrow \infty} \hat{R}_{n} \tag{72}
\end{align*}
$$

It follows from (72) and (13) that

$$
\limsup _{n \rightarrow \infty} E\left(n, \hat{R}_{n}, p\right)=E(R, p)
$$

As $\Delta_{R}>0$, the above equation implies that there exists an $N$ such that for all $n \geq N$, we have

$$
\begin{equation*}
E\left(n, \hat{R}_{n}, p\right)<E(R, p)+\frac{\Delta_{R}}{2} \tag{73}
\end{equation*}
$$

Combining (12), 71, and 73), for $n \geq N$, we obtain

$$
\begin{equation*}
P_{\mathrm{e}}\left(n,\left\lfloor k \varepsilon_{n}\right\rfloor, p\right)>2^{-n\left(E(R, p)+\left(\Delta_{R} / 2\right)\right)} \tag{74}
\end{equation*}
$$

Now, if we define $\beta_{n}$ as

$$
\beta_{n} \triangleq-\frac{1}{n} \log \left((1-\alpha) \varepsilon_{n}\right)
$$

then we have $\beta_{n}>0$, while $\lim _{n \rightarrow \infty} \beta_{n}=0$. Thus, we have

$$
\begin{equation*}
(m-k) \varepsilon_{n}=m(1-\alpha) \varepsilon_{n}=2^{n\left(R-\beta_{n}\right)} \tag{75}
\end{equation*}
$$

Combining (74) and (75), for all $n \geq N$, we have

$$
\begin{align*}
(m-k) \varepsilon_{n} P_{\mathrm{e}}\left(n,\left\lfloor k \varepsilon_{n}\right\rfloor, p\right) & >2^{n\left(R-E(R, p)-\left(\Delta_{R} / 2\right)-\beta_{n}\right)} \\
& =2^{n\left(\left(\Delta_{R} / 2\right)-\beta_{n}\right)} \tag{76}
\end{align*}
$$

As (10) holds for all $0<\varepsilon<1 / 2$ and $k>1 / \varepsilon$, replacing $\varepsilon$ with $\varepsilon_{n}=\frac{1}{n}$ in 10, we get for $n>N$ that

$$
\begin{align*}
& \underline{D}\left(n, 2^{n R}, p, \alpha 2^{n R}\right) \\
& \quad>\left(1-2 \varepsilon_{n}\right)\left[1-\exp \left(-(m-k) \varepsilon_{n} P_{\mathrm{e}}\left(n,\left\lfloor k \varepsilon_{n}\right\rfloor, p\right)\right)\right] \\
& \quad>\left(1-2 \varepsilon_{n}\right)\left[1-\exp \left(-2^{n\left(\left(\Delta_{R} / 2\right)-\beta_{n}\right)}\right)\right] \tag{77}
\end{align*}
$$

where (77) follows from 76. Now, as $\beta_{n}=o(1)$, there exists $\hat{N}$ such that for all $n \geq N$, we have $\Delta_{R} / 2-\beta_{n}>0$. Further, as $\beta_{n}$ is a decreasing function of $n$, it follows that

$$
\begin{equation*}
\lim _{n \rightarrow \infty}\left[1-\exp \left(-2^{n\left(\left(\Delta_{R} / 2\right)-\beta_{n}\right)}\right)\right]=1 \tag{78}
\end{equation*}
$$

As $\lim _{n \rightarrow \infty}\left(1-2 \varepsilon_{n}\right)=1$, combining (77) and 78 with the fact that $\underline{D}\left(n, 2^{n R}, p, \alpha 2^{n R}\right)$ is upper bounded by 1 , we obtain the following important result

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \underline{D}\left(n, 2^{n R}, p, \alpha 2^{n R}\right)=1, \text { for } R>R_{p}^{*} \tag{79}
\end{equation*}
$$

thereby showing that $C_{\underline{D}}(p, \alpha, \epsilon) \leq R_{p}^{*}$, and completing the proof of the strong converse.


Fig. 5: Plot demonstrating $\delta_{\mathrm{GV}}\left(R_{\mathrm{ex}} / 2\right) B_{p}>R_{\mathrm{ex}} / 2$.

## Appendix G <br> Proof of Theorem 3

Proof: We first show that $E(R, p)-R>0$ when $0<$ $R<R_{\mathrm{ex}} / 2$. Towards this, we note from (21) and (22) that $E\left(R_{\mathrm{ex}} / 2, p\right) \geq \delta_{\mathrm{GV}}\left(R_{\mathrm{ex}} / 2\right) B_{p}$. Now, it can be numerically verified that $\delta_{\mathrm{GV}}\left(R_{\mathrm{ex}} / 2\right) B_{p}>\left(R_{\mathrm{ex}} / 2\right)$ when $0<p<0.5$ (see Fig. 5], and hence $E\left(R_{\mathrm{ex}} / 2, p\right)>R_{\mathrm{ex}} / 2$. As $E(R, p)$ is non-increasing in $R$, it follows that

$$
\begin{equation*}
E(R, p)-R>0, \text { when } 0<R<R_{\mathrm{ex}} / 2 \tag{80}
\end{equation*}
$$

Now, for $0<R<\min \left\{0.169, R_{\mathrm{ex}} / 2\right\}$, we have

$$
\begin{aligned}
\lim _{\alpha \downarrow 0} E_{\underline{D}}(R, p, \alpha) & \stackrel{(\mathrm{i})}{=}|E(R, p)-R|^{+} \\
& \stackrel{(\mathrm{ii)}}{=} E(R, p)-R \\
& \stackrel{(\mathrm{iii})}{\leq} \delta_{\mathrm{LP}}(R) B_{p}-R \\
& \stackrel{(\mathrm{iv})}{<} 2 \delta_{\mathrm{GV}}(2 R) B_{p}-R \\
& <2 \delta_{\mathrm{GV}}(2 R) B_{p} \\
& \stackrel{(\mathrm{v})}{\leq} E_{\underline{D}}(R, p)
\end{aligned}
$$

where (i) follows from Thm. 1 and the fact that $E_{\underline{D}}(R, p, \alpha)$ is constant for $0<\alpha<1$, (ii) follows from 80), (iii) follows from (26), (iv) follows from the fact that $\delta_{\mathrm{LP}}(R)<$ $2 \delta_{\mathrm{GV}}(2 R)$ for $0<R<0.169$ (see Fig. 6), and (v) follows from [1]. Theorem 4].

## Acknowledgement

The authors acknowledge discussions with Po-Ning Chen, Fady Alajaji, and Mariam Harutyunyan, on the continuity of the reliability function for discrete memoryless channels.

## REFERENCES

[1] A. Tandon, V. Y. F. Tan, and L. R. Varshney, "The bee-identification problem: Bounds on the error exponent," IEEE Trans. Commun., 2019, [Online early access].


Fig. 6: $\delta_{\mathrm{LP}}(R)<2 \delta_{\mathrm{GV}}(2 R)$ for $0<R<0.169$.
[2] T. Gernat et al., "Automated monitoring of behavior reveals bursty interaction patterns and rapid spreading dynamics in honeybee social networks," Proc. Nat. Acad. Sci. U.S.A., vol. 115, no. 7, pp. 1433-1438, Feb. 2018.
[3] S. Shahi, D. Tuninetti, and N. Devroye, "On identifying a massive number of distributions," in Proc. 2018 IEEE Int. Symp. Inf. Theory, Jun. 2018, pp. 331-335.
[4] A. Pananjady, M. J. Wainwright, and T. A. Courtade, "Linear regression with shuffled data: Statistical and computational limits of permutation recovery," IEEE Trans. Inform. Theory, vol. 64, no. 5, pp. 3286-3300, May 2018.
[5] R. Heckel, I. Shomorony, K. Ramchandran, and D. N. C. Tse, "Fundamental limits of DNA storage systems," in Proc. 2017 IEEE Int. Symp. Inf. Theory, Jun. 2017, pp. 3130-3134.
[6] I. Shomorony and R. Heckel, "Capacity results for the noisy shuffling channel," in Proc. 2019 IEEE Int. Symp. Inf. Theory, Jul. 2019, pp. 762-766.
[7] M. Kovačević and V. Y. F. Tan, "Codes in the space of multisets coding for permutation channels with impairments," IEEE Trans. Inform. Theory, vol. 64, no. 7, pp. 5156-5169, Jul. 2018.
[8] R. G. Gallager, Principles of Digital Communication. Cambridge University Press, 2008.
[9] -, Information Theory and Reliable Communication. New York: John Wiley and Sons, 1968.
[10] T. S. Han, Information-Spectrum Methods in Information Theory, ser. Stochastic Modelling and Applied Probability. Springer, 2003.
[11] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems (2nd ed). Cambridge Univ. Press, 2011.
[12] E. A. Haroutunian, "On the optimality of information transmission by a channel with finite number of states known at the input," Izvestiya Akademii Nauk Armenii, Matematika, vol. 4, no. 2, pp. 81-90, 1969, (in Russian).
[13] E. A. Haroutunian, M. E. Haroutunian, and A. N. Harutyunyan, "Reliability criteria in information theory and in statistical hypothesis testing," Foundations and Trends in Communications and Information Theory, vol. 4, no. 23, pp. 97-263, 2008.
[14] N. Weinberger and N. Merhav, "Optimum tradeoffs between the error exponent and the excess-rate exponent of variable-rate Slepian-Wolf coding," IEEE Trans. Inform. Theory, vol. 61, no. 4, pp. 2165-2190, Apr. 2015.
[15] P.-N. Chen and F. Alajaji, "Optimistic Shannon coding theorems for arbitrary single-user systems," IEEE Trans. Inform. Theory, vol. 45, no. 7, pp. 2623-2629, Nov. 1999.
[16] R. McEliece and J. Omura, "An improved upper bound on the block coding error exponent for binary-input discrete memoryless channels," IEEE Trans. Inform. Theory, vol. 23, no. 5, pp. 611-613, Sep. 1977.
[17] W. Rudin, Principles of Mathematical Analysis, ser. International Series in Pure and Applied Mathematics. McGraw-Hill, 1964.
[18] J. Wolfowitz, Coding Theorems of Information Theory, 3rd ed. Springer-Verlag, New York, 1978.
[19] A. Barg and G. D. Forney, "Random codes: minimum distances and error exponents," IEEE Trans. Inform. Theory, vol. 48, no. 9, pp. 2568-2573, Sep. 2002.
[20] A. Barg and A. McGregor, "Distance distribution of binary codes and the error probability of decoding," IEEE Trans. Inform. Theory, vol. 51, no. 12, pp. 4237-4246, Dec. 2005.
[21] R. McEliece, E. Rodemich, H. Rumsey, and L. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities," IEEE Trans. Inform. Theory, vol. 23, no. 2, pp. 157-166, Mar. 1977.
[22] S. Litsyn, "New upper bounds on error exponents," IEEE Trans. Inform. Theory, vol. 45, no. 2, pp. 385-398, Mar. 1999.
[23] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. II," Inf. Control, vol. 10, no. 5, pp. 522-552, 1967.
[24] N. Merhav, "Universal decoding for arbitrary channels relative to a given class of decoding metrics," IEEE Trans. Inform. Theory, vol. 59, no. 9, pp. 5566-5576, Sep. 2013.
[25] N. Shulman, "Communication over an unknown channel via common broadcasting," PhD dissertation, Tel Aviv Univ., Israel, 2003.
[26] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," Inf. Control, vol. 10, no. 1, pp. 65-103, 1967.


[^0]:    ${ }^{1}$ We will remove the ceiling operator subsequently; this does not affect the asymptotic behavior of the error exponent $-\frac{1}{n} \log \underline{D}(n, m, p, k)$.
    ${ }^{2}$ Another popular, though perhaps pessimistic, definition of the reliability function given by Han [10] and Csiszár-Körner [11], replaces limsup with liminf in 11 .

[^1]:    ${ }^{3}$ We will remove the floor operator subsequently.
    ${ }^{4}$ We remark that the result in Theorem 1 goes through verbatim if we replace lim sup with liminf in definitions 11) and 14 .

[^2]:    ${ }^{5}$ This is analogous to the optimistic $\epsilon$-capacity defined by Chen and Alajaji 15 Def. 4.10].

