Bee-Identification Error Exponent with Absentee Bees

Anshoo Tandon
National University of Singapore
anshoo.tandon@gmail.com

Vincent Y. F. Tan
National University of Singapore
vtan@nus.edu.sg

Lav R. Varshney
University of Illinois at Urbana-Champaign,
and Salesforce Research
varshney@illinois.edu

***Abstract*—The "bee-identification problem" was formally defined by Tandon, Tan and Varshney [*IEEE Trans. Commun.*, vol. 67, 2019], 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.**

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 finite-resolution 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 $\mathcal{C}$ 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 π 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 $\{1,\dots,m-k\}$ to $\{1,\dots,m\}$ and corresponds to the restriction of permutation π 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.

This work was supported in part by a Singapore Ministry of Education Tier 2 grant (R-263-000-C83-112).

Fig. 1: Bees tagged with barcodes (photograph provided by T. Gernat and G. Robinson).

When $j=\pi(i)$ and the j-th row of codebook $\mathcal{C}$ is denoted $\boldsymbol{c}_j=[c_{j,1}\ c_{j,2}\ \cdots\ c_{j,n}]$, then the i-th row of $\mathcal{C}_\pi$ is equal to $\boldsymbol{c}_j$. For $1\le i\le m-k$, 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

$$\Pr\{\tilde{\boldsymbol{c}}_i \mid \boldsymbol{c}_{\pi(i)}\} = p^{d_i}(1-p)^{n-d_i}, \quad 1\le i\le m-k,$$

$$\Pr\left\{\tilde{\mathcal{C}}_{\pi_{(m-k)}} \,\middle|\, \mathcal{C}, \pi_{(m-k)}\right\} = \prod_{i=1}^{m-k} p^{d_i}(1-p)^{n-d_i}, \tag{1}$$

where $d_i \triangleq \mathrm{d_H}(\tilde{\boldsymbol{c}}_i, \boldsymbol{c}_{\pi(i)})$ denotes the Hamming distance between vectors $\tilde{\boldsymbol{c}}_i$ and $\boldsymbol{c}_{\pi(i)}$.

We remark that the bee-identification problem formulation has other applications in engineering, such as package-distribution to recipients from deliveries with noisy address labels, and identification of warehouse products using wide-area

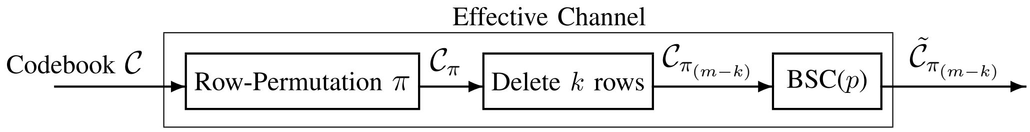

Fig. 2: Effective channel for the bee-identification problem with k absentee bees.

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]. The effective channel in [6] is closely related to the bee-identification channel.

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 is a function ϕ that takes $\tilde{\mathcal{C}}_{\pi_{(m-k)}}$ as an input and produces a map $\nu : \{1, \ldots, m-k\} \to \{1, \ldots, m\}$ where $\nu(i)$ corresponds to the index of the transmitted codeword which produced the received word $\tilde{\boldsymbol{c}}_i$, where $1 \le i \le m-k$. 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 Υ 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 ϕ, the expected bee-identification error probability over the BSC(p) is

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

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 Υ. Note that (2) can be equivalently expressed as

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

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

$$\underline{D}(n, m, p, k) \triangleq \min_{\mathcal{C}, \phi} D(\mathcal{C}, p, k, \phi), \tag{4}$$

where the minimum is over all codebooks $\mathcal{C} \in \mathscr{C}(n, m)$, and all decoding functions ϕ. 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 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 noisy barcodes. This is in contrast to the result *without* absentee bees [1], where joint decoding results in a significantly higher exponent than independent barcode decoding.

Moreover, 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.

Due to space constraints, certain detailed proofs are omitted and we refer readers to the full version of the paper in [7].

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 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 $1 \le i \le m-k$, the decoder picks $\tilde{\boldsymbol{c}}_i$, the i-th row of $\tilde{\mathcal{C}}_{\pi_{(m-k)}}$, and then decodes it to $\nu(i) = \arg\min_{1 \le j \le m} \mathrm{d_H}(\tilde{\boldsymbol{c}}_i, \boldsymbol{c}_j)$. If there is more than one codeword at the same minimum Hamming distance from $\tilde{\boldsymbol{c}}_i$, then any corresponding codeword index is chosen uniformly at random.

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

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

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 transmitting information over BSC(p), let $P_{\rm e}(n,m,p)$ denote the minimum achievable average error probability, where the minimization is over all codebooks $\mathcal{C} \in \mathscr{C}(n,m)$. The following lemma presents an upper bound on $\underline{D}(n,m,p,k)$ in terms of $P_{\rm e}(n,m,p)$.

Lemma 1. *Using independent barcode decoding, the bee-identification error probability $\underline{D}(n,m,p,k)$ can be upper bounded as follows*

$$\underline{D}(n,m,p,k) \leq \min\{1,\, (m-k)\, P_{\rm e}(n,m,p)\}. \quad (6)$$

Proof: Follows from (5), and the fact that

$$\min_{\mathcal{C}\in\mathscr{C}(n,m)} \mathbb{E}_{\pi_{(m-k)}}\left[\Pr\left\{\nu(i)\neq\pi_{(m-k)}(i)\right\}\right] = P_{\rm e}(n,m,p).$$

■

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

Recall Υ 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 ϕ takes the channel output $\tilde{\mathcal{C}}_{\pi_{(m-k)}}$ as an input, and produces the map

$$\nu = \arg\min_{\sigma\in\Upsilon} \mathrm{d_H}(\tilde{\mathcal{C}}_{\pi_{(m-k)}}, \mathcal{C}_\sigma), \quad (7)$$

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 $\mathrm{d_H}(\tilde{\mathcal{C}}_{\pi_{(m-k)}}, \mathcal{C}_\sigma) \triangleq |\{(i,j) : \tilde{\mathcal{C}}_{\pi_{(m-k)}}(i,j) \neq \mathcal{C}_\sigma(i,j), 1\leq i\leq m-k, 1\leq j\leq n\}|$. For this joint ML decoding scheme, we denote the decoding function as $\phi_{\rm J}$. As $\pi_{(m-k)}$ is uniformly distributed over Υ, the joint ML decoder minimizes the error probability [8, Thm. 8.1.1], and from (4) we have

$$\underline{D}(n,m,p,k) = \min_{\mathcal{C}\in\mathscr{C}(n,m)} D(\mathcal{C},p,k,\phi_{\rm J}). \quad (8)$$

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

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

$$\underline{D}(n,m,p,k) > \frac{1-2\varepsilon}{2}\min\left\{1, (m-k)\varepsilon\, P_{\rm e}(n,\lfloor k\varepsilon\rfloor,p)\right\}. \quad (9)$$

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

$$(1-2\varepsilon)\left[1-\exp\left(-(m-k)\varepsilon\, P_{\rm e}(n,\lfloor k\varepsilon\rfloor,p)\right)\right]. \quad (10)$$

Proof (Sketch): The lower bound in (9) follows from Shulman's lower bound on the probability of the union of *pairwise-independent* error events [9, Eq. (30)], [10, p. 109], whereas (10) uses the fact that if $\mathscr{E}^{(\ell)}$, for $1\leq\ell\leq m-k$, are *mutually independent* error events then $\Pr\left\{\bigcup_{1\leq\ell\leq m-k}\mathscr{E}^{(\ell)}\right\} = 1 - \exp\left(\sum_{\ell=1}^{m-k}\ln\left(1-\Pr\left\{\mathscr{E}^{(\ell)}\right\}\right)\right)$. See [7] for a detailed proof. ■

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 $P_{\rm e}(n,m,p)$ denotes the minimum achievable average error probability when m binary codewords, each having blocklength n, are used for transmission of information over BSC(p). For a given $R>0$ and $m=\lceil 2^{nR}\rceil$,[1] the *reliability function* of the channel BSC(p) is defined as follows [11],[2]

$$E(R,p) \triangleq \limsup_{n\to\infty} -\frac{1}{n}\log P_{\rm e}(n,2^{nR},p). \quad (11)$$

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

$$E(n,R_n,p) \triangleq -\frac{1}{n}\log P_{\rm e}(n,2^{nR_n},p). \quad (12)$$

We will relate $E(n,R_n,p)$ to $E(R,p)$ via Lemma 3.

Lemma 3. *Assume that the sequence $(R_n)_{n\in\mathbb{N}}$ converges, and that $R=\lim_{n\to\infty}R_n$. Then we have*

$$\limsup_{n\to\infty} E(n,R_n,p) = E(R,p). \quad (13)$$

Proof (Sketch): Follows from the continuity of the reliability function. See [7] for a detailed proof. ■

Lemma 3 will be pivotal in establishing the exact bee-identification 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 $m=2^{nR}$.
- For a given $0<\alpha<1$, the number of absentee bees k scale as $k=\lfloor\alpha m\rfloor$, where α denotes the fraction of bees missing from the camera image.[3]

For this scenario, define the bee-identification exponent as[4]

$$E_{\underline{D}}(R,p,\alpha) \triangleq \limsup_{n\to\infty} -\frac{1}{n}\log\underline{D}(n,m,p,k). \quad (14)$$

[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 [12] and Csiszár-Körner [13], replaces $\limsup$ with $\liminf$ in (11).

[3] We will assume $k=\alpha m$, and drop the floor operator, subsequently.

[4] We remark that the result in Theorem 1 goes through verbatim if we replace $\limsup$ with $\liminf$ in definitions (11) and (14).

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*

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

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

Proof: We first show that $E_{\underline{D}}(R, p, \alpha) \geq |E(R, p) - R|^+$ using the upper bound on $\underline{D}(n, m, p, k)$ with independent barcode decoding in Lemma 1. Towards this, we note that when $m = 2^{nR}$ and $k = \alpha m$, for $0 < \alpha < 1$, then we have

$$\begin{aligned} \lim_{n\to\infty} -\frac{1}{n}\log(m-k) &= \left(\lim_{n\to\infty} -\frac{1}{n}\log(1-\alpha)\right) - R, \\ &= -R. \end{aligned} \tag{16}$$

Combining (6), (14), and (16), we get

$$\begin{aligned} E_{\underline{D}}(R, p, \alpha) &\geq \left|\left(\limsup_{n\to\infty} -\frac{1}{n}\log P_{\mathrm{e}}(n, m, p)\right) - R\right|^+, \\ &= |E(R, p) - R|^+, \end{aligned} \tag{17}$$

where the last equality follows from (11).

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

$$\hat{R}_n \triangleq \frac{1}{n}\log(\lfloor k\varepsilon \rfloor). \tag{18}$$

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

$$R + \frac{1}{n}\log\left(\frac{\alpha\varepsilon}{2}\right) < \hat{R}_n \leq R + \frac{1}{n}\log(\alpha\varepsilon), \tag{19}$$

which implies that

$$R = \lim_{n\to\infty} \hat{R}_n. \tag{20}$$

Combining the facts that $\lim_{n\to\infty} \frac{1}{n}\log\left((1-2\varepsilon)/2\right) = 0$, $\lim_{n\to\infty} \frac{1}{n}\log\left((m-k)\varepsilon\right) = R$, with (14), (9), (12), and (18), we get

$$\begin{aligned} E_{\underline{D}}(R, p, \alpha) &\leq \left|\limsup_{n\to\infty} E(n, \hat{R}_n, p) - R\right|^+, \\ &= |E(R, p) - R|^+, \end{aligned} \tag{21}$$

where the last equality follows from (20) and (13). The proof is now complete by combining (17) and (21). ■

The above theorem implies the following remarks.

Remark 1. *For a given $0 < \alpha < 1$, if the number of absentee bees k scales as αm, 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 [7]. The proof of Theorem 1 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 $m = 2^{nR}$ with $R > 0$, 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 (α, ϵ)*-achievable* if $\liminf_{n\to\infty} \underline{D}(n, 2^{nR}, p, \alpha 2^{nR}) \leq \epsilon$, and define the ϵ-*capacity* of the bee-identification problem as the supremum of all (α, ϵ)-achievable rates. We denote this ϵ-capacity as[5]

$$C_{\underline{D}}(p, \alpha, \epsilon) \triangleq \sup\left\{R : \liminf_{n\to\infty} \underline{D}(n, 2^{nR}, p, \alpha 2^{nR}) \leq \epsilon\right\}. \tag{22}$$

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

Now, the *Bhattacharyya parameter* for BSC(p) is [15]

$$B_p \triangleq -\log\sqrt{4p(1-p)}, \tag{23}$$

and it is well known that [15]

$$\lim_{R\downarrow 0} E(R, p) = \frac{B_p}{2}. \tag{24}$$

For a given $0 < p < 0.5$, define the function $f(R) \triangleq E(R, p) - R$. From (23) and (24), 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 [7]. Therefore, it follows from the *intermediate value theorem* [16] 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*

$$C_{\underline{D}}(p, \alpha, \epsilon) = R_p^*, \tag{25}$$

where R_p^* is unique positive solution of the following equation

$$E(R, p) = R. \tag{26}$$

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

[5]This is analogous to the *optimistic ϵ-capacity* defined by Chen and Alajaji [14, Def. 4.10].

R_p^* that $E_{\underline{D}}(R,p,\alpha)$ is *strictly* positive. Therefore, it follows from (14) that there exist infinitely many n for which

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

In other words, when $R < R_p^*$, $\underline{D}(n,2^{nR},p,\alpha 2^{nR}) < 2^{-nE_{\underline{D}}(R,p,\alpha)/2}$. Thus when $R < R_p^*$, we have

$$\liminf_{n\to\infty} \underline{D}(n,2^{nR},p,\alpha 2^{nR}) = 0.$$

Therefore, any rate less than R_p^* is achievable and it follows from the definition of capacity in (22) that $C_{\underline{D}}(p,\alpha,\epsilon) \geq R_p^*$.

Next, we will apply the lower bound on $\underline{D}(n,m,p,k)$, given by (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 $\underline{D}(n,2^{nR},p,\alpha 2^{nR})$ tends to 1 as $n \to \infty$. 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 $k = \alpha 2^{nR} > 2n = 2/\varepsilon_n$. Now define $\hat{R}_n$ to be

$$\hat{R}_n \triangleq \frac{1}{n}\log(\lfloor k\varepsilon_n \rfloor). \tag{27}$$

Then, we have

$$R + \frac{1}{n}\log\left(\frac{\alpha}{2n}\right) < \hat{R}_n \leq R + \frac{1}{n}\log\left(\frac{\alpha}{n}\right), \quad R = \lim_{n\to\infty}\hat{R}_n. \tag{28}$$

It follows from (28) and (13) that

$$\limsup_{n\to\infty} E(n,\hat{R}_n,p) = E(R,p).$$

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

$$E(n,\hat{R}_n,p) < E(R,p) + \frac{\Delta_R}{2}. \tag{29}$$

Combining (12), (27), and (29), for $n \geq N$, we obtain

$$P_{\mathrm{e}}(n,\lfloor k\varepsilon_n\rfloor,p) > 2^{-n(E(R,p)+(\Delta_R/2))}. \tag{30}$$

Now, if we define β_n as

$$\beta_n \triangleq -\frac{1}{n}\log\left((1-\alpha)\varepsilon_n\right),$$

then we have $\beta_n > 0$, while $\lim_{n\to\infty}\beta_n = 0$. Thus, we have

$$(m-k)\varepsilon_n = m(1-\alpha)\varepsilon_n = 2^{n(R-\beta_n)}. \tag{31}$$

Combining (30) and (31), for all $n \geq N$, we have

$$(m-k)\varepsilon_n P_{\mathrm{e}}(n,\lfloor k\varepsilon_n\rfloor,p) > 2^{n(R-E(R,p)-(\Delta_R/2)-\beta_n)}, \quad = 2^{n((\Delta_R/2)-\beta_n)}. \tag{32}$$

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

$$\begin{aligned}&\underline{D}(n,2^{nR},p,\alpha 2^{nR})\\ &> (1-2\varepsilon_n)\left[1-\exp\left(-(m-k)\varepsilon_n P_{\mathrm{e}}(n,\lfloor k\varepsilon_n\rfloor,p)\right)\right],\\ &> (1-2\varepsilon_n)\left[1-\exp\left(-2^{n((\Delta_R/2)-\beta_n)}\right)\right],\end{aligned} \tag{33}$$

where (33) follows from (32). Now, as $\beta_n = o(1)$, there exists $\hat{N}$ such that for all $n \geq \hat{N}$, we have $\Delta_R/2 - \beta_n > 0$. Further, as β_n is a decreasing function of n, it follows that

$$\lim_{n\to\infty}\left[1-\exp\left(-2^{n((\Delta_R/2)-\beta_n)}\right)\right] = 1. \tag{34}$$

As $\lim_{n\to\infty}(1-2\varepsilon_n) = 1$, combining (33) and (34) with the fact that $\underline{D}(n,2^{nR},p,\alpha 2^{nR})$ is upper bounded by 1, we obtain the following important result

$$\lim_{n\to\infty}\underline{D}(n,2^{nR},p,\alpha 2^{nR}) = 1, \text{ for } R > R_p^*, \tag{35}$$

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

Note that the above proof shows that if $R > R_p^*$, then the error probability $\underline{D}(n,2^{nR},p,\alpha 2^{nR})$ tends to 1 as $n \to \infty$, thereby proving the *strong converse property* [17]. Also note that the expression for the ϵ-capacity (25) is independent of the value of $\alpha \in (0,1)$, and that a similar behavior was observed for the bee-identification exponent (15).

IV. Reflections

This work extended the characterization of the bee-identification 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* [11].

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 α 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 bee-identification 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 α when $0 < \alpha < 1$. We also characterized the *exact* 'capacity' for the bee-identification problem with absentee bees, and proved the *strong converse*.

Future work includes exploring the error exponent for the scenario where α, 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.

References

[1] A. Tandon, V. Y. F. Tan, and L. R. Varshney, "The bee-identification problem: Bounds on the error exponent," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7405–7416, Nov. 2019.

[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] A. Tandon, V. Y. F. Tan, and L. R. Varshney, “On the bee-identification error exponent with absentee bees,” Oct. 2019, arXiv:1910.10333 [cs.IT].
[8] R. G. Gallager, *Principles of Digital Communication*. Cambridge University Press, 2008.
[9] 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.
[10] N. Shulman, “Communication over an unknown channel via common broadcasting,” PhD dissertation, Tel Aviv Univ., Israel, 2003.
[11] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, 1968.
[12] T. S. Han, *Information-Spectrum Methods in Information Theory*, ser. Stochastic Modelling and Applied Probability. Springer, 2003.
[13] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems (2nd ed)*. Cambridge Univ. Press, 2011.
[14] 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.
[15] 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.
[16] W. Rudin, *Principles of Mathematical Analysis*, ser. International Series in Pure and Applied Mathematics. McGraw-Hill, 1964.
[17] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. Springer-Verlag, New York, 1978.