The Bee-Identification Problem: Bounds on the Error Exponent

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

***Abstract*— Consider the problem of identifying a massive number of bees, uniquely labeled with barcodes, using noisy measurements. We formally introduce this "bee-identification problem", define its error exponent, and derive efficiently computable upper and lower bounds for this exponent. We show that joint decoding of barcodes provides a significantly better exponent compared to separate decoding followed by permutation inference. For low rates, we prove that the lower bound on the bee-identification exponent obtained using typical random codes (TRC) is strictly better than the corresponding bound obtained using a random code ensemble (RCE). Further, as the rate approaches zero, we prove that the upper bound on the bee-identification exponent meets the lower bound obtained using TRC with joint barcode decoding.**

***Index Terms*— Bee-identification problem, error exponent, noisy channel, joint decoding, permutation recovery.**

I. Introduction

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]. Assume that a camera is employed to picture the beehive to study the interactions among bees. The image output (see Fig. 1) can be considered as a noisy and unordered set of m barcodes. We formally pose the problem of bee-identification from a beehive image as an information-theoretic problem (Sec. I-B).

The bee-identification problem has applications in identification of warehouse products (labeled with unique RFID barcodes) using wide-area sensors. Other applications include package-distribution to recipients from a batch of deliveries with noisy address labels, and similar "bipartite matching" settings. It also has potential applications in identification of the mapping between signals and their meaning in "alien communication" with extraterrestrials, and also in learning communication protocols among robots, via the use of pilot signals going through the alphabet.

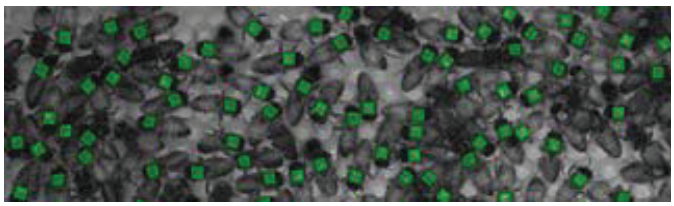

Fig. 1. Bees tagged with barcodes (adapted from [1]).

We consider the scenario where the barcode for each bee is represented as a binary vector of length n, and the bee barcodes are collected in a codebook 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 rows of C with a random permutation π to produce C_π. The entries of C_π are then subjected to noise (corresponding to a binary symmetric channel (BSC) with crossover probability p), and the channel output is denoted $\tilde{C}_\pi$. We assume that the decoder has knowledge of codebook C, and its task is to *recover the row-permutation* π introduced by the channel. Note that the permutation π directly ascertains the identity of all the bees.

A. Related Work

In a related work motivated by an Internet of Things (IoT) setting, the identification of users in strongly asynchronous massive access channels was studied [2]. The identification of the underlying distributions of a set of observed sequences (where each sequence is generated i.i.d. by a distinct distribution) was analyzed in [3]. The bee-identification problem, on the other hand, allows codebooks where all barcode sequences are generated using the same underlying distribution. Note that both the bee-identification problem and the distribution identification problem in [3] can be equivalently viewed as permutation recovery problems. Other applications and models in different settings, where permutation recovery arises naturally, are discussed in [4].

In another related work [5], the fundamental limits of data storage via unordered DNA molecules was investigated. Here, a DNA molecule corresponds to an ℓ-length sequence over

Manuscript received May 20, 2019; revised August 1, 2019; accepted August 7, 2019. Date of publication August 13, 2019; date of current version November 19, 2019. This work was supported in part by a Singapore Ministry of Education Tier 2 grant (R-263-000-C83-112) and in part by the National Science Foundation grant CCF-1717530. This article was presented in part at the 2019 IEEE Information Theory Workshop in Visby, Gotland, Sweden. The associate editor coordinating the review of this article and approving it for publication was R. Venkataramanan. *(Corresponding author: Anshoo Tandon.)*

A. Tandon is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583 (e-mail: anshoo.tandon@gmail.com).

V. Y. F. Tan is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583, and also with the Department of Mathematics, National University of Singapore, Singapore 119076 (e-mail: vtan@nus.edu.sg).

L. R. Varshney is with the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Urbana, IL 61801 USA, and also with the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, Urbana, IL 61801 USA (e-mail: varshney@illinois.edu).

Digital Object Identifier 10.1109/TCOMM.2019.2935204

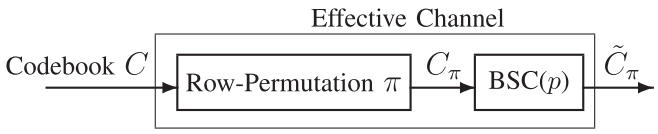

Fig. 2. Effective channel for the bee-identification problem.

an alphabet of size 4, and the information is written onto m DNA molecules stored in an unordered way. The storage capacity results in [5] were extended to noisy settings in [6] where the channel adds noise and randomly permutes the m DNA molecules used to store information. The capacity results are obtained under the scenario where the length, ℓ, of each DNA molecule grows with m. Although the effective channel in [6] is closely related to the bee-identification channel in Fig. 2, we note that the fundamental problem in [6] is to quantify the data storage capacity, while the main issue in the bee-identification problem is the identification of the row-permutation induced by the channel.

Data communication over permutation channels with impairments was analyzed in [7]. The authors of [7] presented bounds on the size of optimal codes over a finite input alphabet, when the channel randomly permutes the letters of the input sequence in addition to causing impairments such as insertions, deletions, and substitutions. The effective channel for the bee-identification problem (see Fig. 2) 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, not 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.

B. Bee-Identification Problem Formulation

The channel output is a row-permuted and noisy version of the codebook. If π denotes a given permutation of m-letters, then the channel first permutes the m rows of codebook C, based on π, to produce C_π (see Fig. 2). Therefore, if $j = \pi(i)$ and the i-th row of codebook C is denoted $\boldsymbol{c}_i = [c_{i,1}\ c_{i,2}\ \cdots\ c_{i,n}]$, then the j-th row of C_π is equal to $\boldsymbol{c}_i$. The channel then applies noise on the permuted codebook C_π to produce $\tilde{C}_\pi$, where noise is modeled by a BSC with crossover probability p, denoted BSC(p), with $0 < p < 0.5$. If $j = \pi(i)$, and $\tilde{\boldsymbol{c}}_{\pi(i)}$ denotes the j-th row of $\tilde{C}_\pi$, then

$$\begin{aligned} &\Pr\{\tilde{\boldsymbol{c}}_{\pi(i)}|\boldsymbol{c}_i,\pi\} = p^{d_i}(1-p)^{n-d_i}, \quad 1 \le i \le m, \\ &\Pr\left\{\tilde{C}_\pi|C,\pi\right\} = \prod_{i=1}^{m}\Pr\{\tilde{\boldsymbol{c}}_{\pi(i)}|\boldsymbol{c}_i,\pi\} = \prod_{i=1}^{m} p^{d_i}(1-p)^{n-d_i}, \end{aligned} \tag{1}$$

where $d_i \triangleq \mathrm{d_H}(\tilde{\boldsymbol{c}}_{\pi(i)}, \boldsymbol{c}_i)$ denotes the Hamming distance between vectors $\tilde{\boldsymbol{c}}_{\pi(i)}$ and $\boldsymbol{c}_i$. Let $\mathcal{M} \triangleq \{1, 2, \ldots, m\}$, and let the decoder correspond to a function ϕ which takes $\tilde{C}_\pi$ as an input and produces a map $\nu : \mathcal{M} \to \mathcal{M}$ where $\nu(k)$ corresponds to the index of the transmitted codeword which produced the received word $\tilde{\boldsymbol{c}}_k$, for $1 \le k \le m$. In effect, the bee-identification problem is that the decoder has to recover the row-permutation π introduced by the channel, by using the knowledge of codebook C and the channel output $\tilde{C}_\pi$.

C. Bee-Identification Error Exponent

The indicator for the bee-identification error is defined as

$$\mathcal{D}\left(\phi(\tilde{C}_\pi), \pi^{-1}\right) = \mathcal{D}\left(\nu, \pi^{-1}\right) \triangleq \begin{cases} 1, & \text{if } \nu \neq \pi^{-1}, \\ 0, & \text{if } \nu = \pi^{-1}. \end{cases}$$

For a given codebook C and decoding function ϕ, the expected bee-identification error probability over the BSC(p) is

$$D(C, p, \phi) \triangleq \mathbb{E}_\pi\left[\mathbb{E}\left[\mathcal{D}\left(\phi(\tilde{C}_\pi), \pi^{-1}\right)\right]\right], \tag{2}$$

where the inner expectation is over the distribution of $\tilde{C}_\pi$ given C and π (see (1)), and the outer expectation is over a uniform distribution of π over all m-letter permutations. Note that (2) can be equivalently expressed as

$$D(C, p, \phi) = \Pr\left\{\phi(\tilde{C}_\pi) \neq \pi^{-1}\right\} = \Pr\left\{\nu \neq \pi^{-1}\right\}. \tag{3}$$

For a given $R > 0$, let the number of barcodes m scale exponentially with blocklength n as $m = 2^{nR}$. Now, for given values of n and R, define the minimum expected bee-identification error probability as

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

where the minimum is over all codebooks C of size $2^{nR} \times n$, and all decoding functions ϕ.

Define, $E_{\underline{D}}(R, p)$, the exponent corresponding to the minimum expected bee-identification error probability, as

$$E_{\underline{D}}(R, p) = \liminf_{n\to\infty} -\frac{1}{n}\log \underline{D}(n, R, p). \tag{5}$$

We introduce some notation that is used in the rest of the paper. We will denote $f(n) \doteq g(n)$ when $\lim_{n\to\infty} n^{-1}\log\left(f(n)/g(n)\right) = 0$. Similarly, we write $f(n) \dot{\le} g(n)$ (respectively, $f(n) \dot{\ge} g(n)$) if $\limsup_{n\to\infty} n^{-1}\log\left(f(n)/g(n)\right) \le 0$ (respectively, ≥ 0). Unless stated otherwise, we will take all logarithms to base 2.

D. Our Contributions

The "bee-identification problem" is introduced and the corresponding bee-identification exponent $E_{\underline{D}}(R, p)$ is analyzed in this paper. In particular, we provide the following *explicit* bounds on this exponent.

- A lower bound on $E_{\underline{D}}(R, p)$ using a random code ensemble (RCE) with independent barcode decoding (Sec. II-A) and joint barcode decoding (Sec. II-B).
- A lower bound on $E_{\underline{D}}(R, p)$ using typical random codes (TRC) with independent barcode decoding (Sec. III-A) and joint barcode decoding (Sec. III-B).
- An upper bound on $E_{\underline{D}}(R, p)$ which is applicable to all possible codebook designs (Sec. IV).

We show that joint decoding of barcodes provides a significantly better exponent compared to separate decoding followed by decoding the permutation. For low rates, we prove that the

lower bound obtained using TRC is strictly better than the corresponding bound obtained using RCE. Further, as the rate approaches zero, we prove that the upper bound meets the lower bound obtained using TRC with joint barcode decoding.

II. Random Code Ensemble

In this section, we present lower bounds on $E_{\underline{D}}(R,p)$ using an RCE [8]. Let $\mathscr{C}(n,R)$ denote the set of all binary matrices with $m=2^{nR}$ rows and n columns. Assume that codebook C is uniformly distributed over $\mathscr{C}(n,R)$. It is immediate from the definition of $\underline{D}(n,R,p)$ (4) that

$$\underline{D}(n,R,p) \leq \frac{1}{|\mathscr{C}(n,R)|}\sum_{C\in\mathscr{C}(n,R)} D(C,p,\phi), \tag{6}$$

where the expression on the right denotes the average performance using RCE. We proceed by quantifying this expression when the decoding function ϕ corresponds to: (i) independent barcode decoding (Sec. II-A), and (ii) joint barcode decoding (Sec. II-B). The main results in this section are as follows: we present explicit lower bounds on $E_{\underline{D}}(R,p)$ using independent barcode decoding (Thm. 1) and joint barcode decoding (Thm. 2). It is shown (Prop. 2) that the bee-identification exponent obtained using joint barcode decoding is strictly better than the corresponding exponent obtained with independent barcode decoding.

A. Independent Decoding for Each Barcode

Here, we analyze a naïve decoding strategy where each barcode is decoded independently. In this case, for $1\leq j\leq m$, the decoder picks $\tilde{\boldsymbol{c}}_j$, the j-th row of $\tilde{C}_\pi$, and then decodes it to $\nu(j)=\arg\min_k \mathrm{d_H}(\tilde{\boldsymbol{c}}_j,\boldsymbol{c}_k)$. If there is more than one codeword at the same minimum Hamming distance from $\tilde{\boldsymbol{c}}_j$, then any one of the corresponding codeword indices is chosen at random. From (3) and the union bound, we have

$$D(C,p,\phi)\leq \sum_{j=1}^{m}\Pr\left\{\nu(j)\neq\pi^{-1}(j)\right\}. \tag{7}$$

Combining (6) and (7), we get

$$\underline{D}(n,R,p)\leq \sum_{j=1}^{m}\left(\sum_{C\in\mathscr{C}(n,R)}\frac{\Pr\left\{\nu(j)\neq\pi^{-1}(j)\right\}}{|\mathscr{C}(n,R)|}\right). \tag{8}$$

Now define

$$P(n,R,p)\triangleq\frac{1}{|\mathscr{C}(n,R)|}\sum_{C\in\mathscr{C}(n,R)}\Pr\left\{\nu(j)\neq\pi^{-1}(j)\right\}. \tag{9}$$

Note that $P(n,R,p)$ is independent of index j due to the averaging over the ensemble of codebooks uniformly distributed over $\mathscr{C}(n,R)$. For $i=\pi^{-1}(j)$, the expression for $P(n,R,p)$ corresponds to the probability of error when the i-th codeword is transmitted over BSC(p). From (8) and (9), we get

$$\underline{D}(n,R,p)\leq mP(n,R,p).$$

Further, the bee-identification error probability $\underline{D}(n,R,p)$ is upper bounded by 1, and so

$$\underline{D}(n,R,p)\leq \min\left\{1,\, mP(n,R,p)\right\}. \tag{10}$$

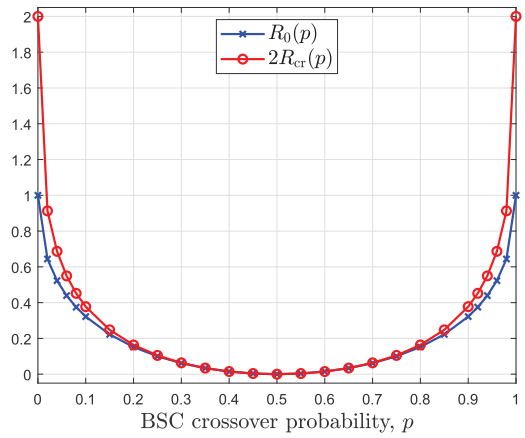

Fig. 3. Plot demonstrating $R_0(p)\leq 2R_{\mathrm{cr}}(p)$.

The following theorem uses (10) to present an explicit lower bound on $E_{\underline{D}}(R,p)$.

Theorem 1: We have

$$E_{\underline{D}}(R,p)\geq |R_0(p)-2R|^+, \tag{11}$$

where $|x|^+\triangleq\max(0,x)$, *and*

$$R_0(p)\triangleq 1-\log\left(1+\sqrt{4p(1-p)}\right). \tag{12}$$

Proof: It is well known that the random coding exponent over BSC(p), defined as $E_{\mathrm{r}}(R,p)\triangleq \liminf_{n\to\infty}(1/n)\log\left(1/P(n,R,p)\right)$, is given by [8], [9]

$$E_{\mathrm{r}}(R,p)=\begin{cases}R_0(p)-R, & 0<R\leq R_{\mathrm{cr}}(p)\\ D(\delta_{\mathrm{GV}}(R)\|p), & R_{\mathrm{cr}}(p)\leq R\leq 1-H(p)\\ 0, & R\geq 1-H(p),\end{cases} \tag{13}$$

where $H(\cdot)$ denotes the binary entropy function, $\delta_{\mathrm{GV}}(R)$ is the Gilbert-Varshamov (GV) distance [8] defined as the value of δ in the interval $[0,0.5]$ with $H(\delta)=1-R$, and $R_{\mathrm{cr}}(p)$ is the critical rate given by $R_{\mathrm{cr}}(p)=1-H\left(\frac{\sqrt{p}}{\sqrt{p}+\sqrt{1-p}}\right)$, and

$$D(x\|y)\triangleq x\log\frac{x}{y}+(1-x)\log\frac{1-x}{1-y}.$$

Using the fact that $m=2^{nR}$, and combining (5), (10), and the definition of $E_{\mathrm{r}}(R,p)$, we get

$$E_{\underline{D}}(R,p)\geq |E_{\mathrm{r}}(R,p)-R|^+. \tag{14}$$

Using explicit numerical computation, it can be shown that $R_0(p)\leq 2R_{\mathrm{cr}}(p)$ (see Fig. 3). The proof is now complete by combining (14) with the first clause of (13), and noting that $|E_{\mathrm{r}}(R,p)-R|^+=0$ when $R\geq R_{\mathrm{cr}}(p)$ as $E_{\mathrm{r}}(R,p)$ is a decreasing function of R. ∎

The lower bound on $E_{\underline{D}}(R,p)$ given by (11) was obtained by applying a naïve decoding strategy where each barcode was decoded independently. In the next subsection, we analyze the bee-identification exponent using joint barcode decoding.

B. Joint Decoding of Barcodes

Let S_m denote the set of permutations of $\{1,\ldots,m\}$. For joint maximum likelihood (ML) decoding of barcodes, the decoding function ϕ takes the noisy row-permuted codebook $\tilde{C}_\pi$ as input, and produces permutation $\nu=\rho^{-1}$ as output, where $\rho=\arg\min_{\sigma\in S_m}\mathrm{d_H}(\tilde{C}_\pi,C_\sigma)$, and $\mathrm{d_H}(\tilde{C}_\pi,C_\sigma)\triangleq|\{(i,j):\tilde{C}_\pi(i,j)\neq C_\sigma(i,j),1\leq i\leq m,1\leq j\leq n\}|$. We aim to provide bounds on $\Pr\{\nu\neq\pi^{-1}\}=\Pr\{\rho\neq\pi\}$.

For any two permutations $\pi_1,\pi_2\in S_m$, the sets of distances $\{\mathrm{d_H}(\tilde{C}_{\pi_1},C_\sigma)\}_{\sigma\in S_m}$ and $\{\mathrm{d_H}(\tilde{C}_{\pi_2},C_\sigma)\}_{\sigma\in S_m}$ are equal. Therefore, the performance of the joint ML decoder is independent of the channel permutation π, and we assume, without loss of generality, that the *permutation induced by the channel is the identity permutation, denoted* π_0.

For a given codebook C at the transmitter, let $\tilde{C}_{\pi_0}$ denote the received noisy codebook at the output of the effective channel, and for $\sigma\in S_m$ with $\sigma\neq\pi_0$, we define

$$\Pr\{\pi_0\to\sigma\}\triangleq\Pr\left\{\mathrm{d_H}(\tilde{C}_{\pi_0},C_\sigma)\leq\mathrm{d_H}(\tilde{C}_{\pi_0},C_{\pi_0})\right\},$$

where the event $\{\pi_0\to\sigma\}$ is said to occur if $\mathrm{d_H}(\tilde{C}_{\pi_0},C_\sigma)\leq\mathrm{d_H}(\tilde{C}_{\pi_0},C_{\pi_0})$. From (3), we have

$$\begin{aligned}D(C,p,\phi)&=\Pr\left\{\bigcup_{\sigma\in S_m,\sigma\neq\pi_0}\{\pi_0\to\sigma\}\right\},\\&\leq\sum_{\sigma\in S_m,\sigma\neq\pi_0}\Pr\{\pi_0\to\sigma\},\end{aligned}\tag{15}$$

where (15) follows from the union bound. Now define

$$P_{\mathrm{RCE},\sigma}\triangleq\frac{1}{|\mathscr{C}(n,R)|}\sum_{C\in\mathscr{C}(n,R)}\Pr\{\pi_0\to\sigma\},\tag{16}$$

which denotes the probability of the event $\{\pi_0\to\sigma\}$, averaged over the ensemble of random binary codebooks. Using (6), (15), and (16), we get

$$\underline{D}(n,R,p)\leq\sum_{\sigma\in S_m,\sigma\neq\pi_0}P_{\mathrm{RCE},\sigma}.\tag{17}$$

Now consider two codewords $\boldsymbol{c}_i$, $\boldsymbol{c}_j$ at distance d from each other. Given that $\boldsymbol{c}_i$ is transmitted over BSC(p), the probability that the Hamming distance of the received word from $\boldsymbol{c}_j$ is not more than its distance from $\boldsymbol{c}_i$ is [8]

$$\Pr\{\boldsymbol{c}_i\to\boldsymbol{c}_j\}\leq 2^{-d\,\alpha_p},$$

where

$$\alpha_p\triangleq-\log\sqrt{4p(1-p)}.\tag{18}$$

Therefore, for a given codebook $C=C_{\pi_0}$ and permutation $\sigma\in S_m$ with $\sigma\neq\pi_0$, if $d_\sigma\triangleq\mathrm{d_H}(C_{\pi_0},C_\sigma)$, then

$$\Pr\{\pi_0\to\sigma\}\leq 2^{-d_\sigma\alpha_p}.\tag{19}$$

In the following, we quantify $P_{\mathrm{RCE},\sigma}$ for different $\sigma\in S_m$, via (16) and (19).

1) σ is a Transposition: We first consider the case where σ is a *transposition*, i.e. a permutation that interchanges only two indices. For indices $\hat{\imath},\hat{\jmath}$, with $1\leq\hat{\imath}<\hat{\jmath}\leq m$, the Hamming distance between codewords $\boldsymbol{c}_{\hat{\imath}}$ and $\boldsymbol{c}_{\hat{\jmath}}$ in a random codebook satisfies [8]

$$\Pr\{\mathrm{d_H}(\boldsymbol{c}_{\hat{\imath}},\boldsymbol{c}_{\hat{\jmath}})=d\}\leq 2^{-n(1-H(d/n))}.\tag{20}$$

When $\sigma=(\hat{\imath}\ \hat{\jmath})$ is the permutation that only transposes indices $\hat{\imath}$ and $\hat{\jmath}$, then $\mathrm{d_H}\left(C_{\pi_0},C_{(\hat{\imath}\ \hat{\jmath})}\right)=2d$ if and only if $\mathrm{d_H}(\boldsymbol{c}_{\hat{\imath}},\boldsymbol{c}_{\hat{\jmath}})=d$. Thus, it follows from (20) that $\Pr\left\{\mathrm{d_H}\left(C_{\pi_0},C_{(\hat{\imath}\ \hat{\jmath})}\right)=2d\right\}\leq 2^{-n(1-H(d/n))}$. Further, when $\mathrm{d_H}\left(C_{\pi_0},C_{(\hat{\imath}\ \hat{\jmath})}\right)=2d$, we have $\Pr\{\pi_0\to(\hat{\imath}\ \hat{\jmath})\}\leq 2^{-2d\,\alpha_p}$. Therefore, the probability $P_{\mathrm{RCE},(\hat{\imath}\ \hat{\jmath})}$ can be characterized using (16), (19), and (20) as

$$P_{\mathrm{RCE},(\hat{\imath}\ \hat{\jmath})}\leq\sum_{d=0}^{n}2^{-n(1-H(d/n)+2(d/n)\alpha_p)}.\tag{21}$$

If $\delta=d/n$ is treated as a continuous variable, then the exponent $E_2(\delta)\triangleq 1-H(\delta)+2\delta\alpha_p$ is a convex function with a unique minimum at $\delta=\hat{\delta}_p$ where

$$\hat{\delta}_p\triangleq\frac{4p(1-p)}{1+4p(1-p)}.\tag{22}$$

Therefore, for $0\leq d\leq n$, we have

$$2^{-n(1-H(d/n)+2(d/n)\alpha_p)}\leq 2^{-n\left(1-H(\hat{\delta}_p)+2(\hat{\delta}_p)\alpha_p\right)}.$$

Now, if we define

$$c_n\triangleq\left(\log(n+1)\right)/n,$$

then it follows from (21) that

$$P_{\mathrm{RCE},\sigma}\leq 2^{-n\left(1-H(\hat{\delta}_p)+2(\hat{\delta}_p)\alpha_p-c_n\right)}.\tag{23}$$

Further, we have $1-H(\hat{\delta}_p)+2(\hat{\delta}_p)\alpha_p=R_1(p)$, where

$$R_1(p)\triangleq 1-\log(1+4p(1-p)).\tag{24}$$

Hence, it follows from (23) and (24) that

$$P_{\mathrm{RCE},\sigma}\leq 2^{-n(R_1(p)-c_n)},\tag{25}$$

where σ is a transposition.

2) σ is a Product (Composition) of Disjoint Transpositions: We now consider the case where $\sigma=\sigma_1\sigma_2$, where σ_1 and σ_2 are disjoint transpositions with $\sigma_1=(i\ j)$ and $\sigma_2=(\hat{\imath}\ \hat{\jmath})$. As the codewords in a random codebook are independent, then using (20), we have $\Pr\{\{\mathrm{d_H}(\boldsymbol{c}_i,\boldsymbol{c}_j)=d_1\}\cap\{\mathrm{d_H}(\boldsymbol{c}_{\hat{\imath}},\boldsymbol{c}_{\hat{\jmath}})=d_2\}\}\leq\prod_{i=1}^{2}2^{-n(1-H(d_i/n))}$. Further, if $\mathrm{d_H}(\boldsymbol{c}_i,\boldsymbol{c}_j)=d_1$ and $\mathrm{d_H}(\boldsymbol{c}_{\hat{\imath}},\boldsymbol{c}_{\hat{\jmath}})=d_2$, then $\mathrm{d_H}\left(C_{\pi_0},C_\sigma\right)=2(d_1+d_2)$, and $\Pr\{\pi_0\to\sigma\}\leq 2^{-2(d_1+d_2)\alpha_p}$. Therefore, if σ is a product of two disjoint transpositions, then

$$\begin{aligned}P_{\mathrm{RCE},\sigma}&\leq\sum_{\substack{0\leq d_1\leq n,\\0\leq d_2\leq n}}2^{-n\left(\sum_{i=1}^{2}(1-H(d_i/n)+2(d_i/n)\alpha_p)\right)},\\&=\prod_{i=1}^{2}\left(\sum_{d_i=0}^{n}2^{-n(1-H(d_i/n)+2(d_i/n)\alpha_p)}\right),\\&\leq 2^{-2n(R_1(p)-c_n)}.\end{aligned}$$

In general, when σ is a product of s disjoint transpositions, the above argument can be readily extended to show that

$$P_{\mathrm{RCE},\sigma} \leq 2^{-sn(R_1(p)-c_n)}. \tag{26}$$

Now, define

$$\lambda_p \triangleq \min\left\{\frac{2R_0(p)}{3}, \frac{R_1(p)}{2}\right\},$$

where $R_0(p)$ and $R_1(p)$ are defined in (12) and (24), respectively. As $2\lambda_p \leq R_1(p)$, it follows from (26) that

$$P_{\mathrm{RCE},\sigma} \leq 2^{-n2s(\lambda_p-c_n)}. \tag{27}$$

We remark that when σ is just a transposition, then from (25) we have $P_{\mathrm{RCE},\sigma} \leq 2^{-n(R_1(p)-c_n)} \leq 2^{-n2(\lambda_p-c_n)}$, which is only a special case of (27) with $s=1$.

3) σ is a k-Cycle With $k>2$: Let $\sigma \in S_m$ be a k-cycle $(i_1\ i_2\ \cdots\ i_k)$ where $i_{l+1}=\sigma(i_l)$ for $1 \leq l \leq k-1$, and $i_1=\sigma(i_k)$. We will apply the following proposition towards characterizing $P_{\mathrm{RCE},\sigma}$.

Proposition 1: Let $\boldsymbol{c}_{i_1}, \boldsymbol{c}_{i_2}, \ldots, \boldsymbol{c}_{i_k}$ be k distinct rows in the codebook C, and let d_l satisfy $0 \leq d_l \leq n$ for $1 \leq l \leq k-1$. When C is uniformly distributed over $\mathscr{C}(n,R)$, then the following inequality holds

$$\Pr\left\{\bigcap_{l=1}^{k-1}\left\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}})=d_l\right\}\right\} \leq \prod_{l=1}^{k-1} 2^{-n(1-H(d_l/n))}. \tag{28}$$

Proof: See Appendix A. ∎

For a given codebook C, if $\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l$ for $1 \leq l \leq k-1$, and $\mathrm{d_H}(\boldsymbol{c}_{i_k}, \boldsymbol{c}_{i_1}) = d_k$, then $\mathrm{d_H}(C_{\pi_0}, C_\sigma) = \sum_{l=1}^{k} d_l$, and we have

$$\Pr\{\pi_0 \to \sigma\} \leq 2^{-\left(\sum_{l=1}^{k} d_l\right)\alpha_p}. \tag{29}$$

Further, if codebook C is uniformly distributed over $\mathscr{C}(n,R)$,

$$\Pr\left\{\left(\bigcap_{l=1}^{k-1}\left\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\right\}\right)\bigcap\left\{\mathrm{d_H}(\boldsymbol{c}_{i_k}, \boldsymbol{c}_{i_1}) = d_k\right\}\right\} \leq 2^{-n\left(\sum_{l=1}^{k-1}(1-H(d_l/n))\right)}, \tag{30}$$

where (30) follows from (28). Combining (29) and (30),

$$\begin{aligned} P_{\mathrm{RCE},\sigma} &\leq \sum_{\substack{0\leq d_l \leq n,\\ 1\leq l\leq k}} 2^{-n\left(\left(\sum_{l=1}^{k}(d_l/n)\alpha_p\right)+\left(\sum_{l=1}^{k-1}(1-H(d_l/n))\right)\right)}, \\ &= \sum_{d_k=0}^{n} 2^{-d_k\alpha_p}\left(\prod_{l=1}^{k-1}\sum_{d_l=0}^{n} 2^{-n(1-H(d_l/n)+(d_l/n)\alpha_p)}\right) \\ &\leq 2^{nc_n}\left(\prod_{l=1}^{k-1}\sum_{d_l=0}^{n} 2^{-n(1-H(d_l/n)+(d_l/n)\alpha_p)}\right), \end{aligned} \tag{31}$$

If $\delta = d_l/n$ is treated as a continuous variable, then the exponent $E_1(\delta) \triangleq 1-H(\delta)+\delta\alpha_p$ is a convex function with a unique minimum at $\delta=\tilde{\delta}_p$, where

$$\tilde{\delta}_p \triangleq \frac{\sqrt{4p(1-p)}}{1+\sqrt{4p(1-p)}}. \tag{32}$$

We have

$$E_1(\tilde{\delta}_p) = 1-\log(1+\sqrt{4p(1-p)}) = R_0(p),$$

and therefore

$$\sum_{d_l=0}^{n} 2^{-n(1-H(d_l/n)+(d_l/n)\alpha_p)} \leq 2^{-n(R_0(p)-c_n)}. \tag{33}$$

Combining (31) and (33),

$$P_{\mathrm{RCE},\sigma} \leq 2^{-n((k-1)R_0(p)-kc_n)}. \tag{34}$$

As $2k/3 \leq k-1$ for $k>2$, we have $k\lambda_p \leq 2kR_0(p)/3 \leq (k-1)R_0(p)$, and it follows from (34) that

$$P_{\mathrm{RCE},\sigma} \leq 2^{-nk(\lambda_p-c_n)}. \tag{35}$$

The above equation has been derived for the case where σ is a k-cycle with $k>2$. However, a transposition is just a k-cycle with $k=2$, and from the remark following (27), it follows that (35) holds even for $k=2$.

4) General $\sigma \in S_m$ With $\sigma \neq \pi_0$: It is well known that any permutation $\sigma \neq \pi_0$ can be written as a product (composition) of t disjoint cycles, for $t \geq 1$ [10]. Consider a given σ which is a product of t disjoint cycles of length $k_1, \ldots, k_t$, respectively, where $k_i \geq 2$ for $1 \leq i \leq t$. Then, we can extend the result in (35) to obtain

$$P_{\mathrm{RCE},\sigma} \leq 2^{-n\left(\sum_{i=1}^{t} k_i\right)(\lambda_p-c_n)}. \tag{36}$$

5) Putting it all Together: For $1 \leq j \leq m$, if we define

$$\Sigma_j \triangleq \{\sigma \in S_m : |\{i : \sigma(i) \neq i, 1 \leq i \leq m\}| = j\}, \tag{37}$$

$$P_{\mathrm{RCE},\Sigma_j} \triangleq \sum_{\sigma\in\Sigma_j} P_{\mathrm{RCE},\sigma}\ , \tag{38}$$

then (17) can be equivalently expressed as

$$\underline{D}(n,R,p) \leq \sum_{j=2}^{m} P_{\mathrm{RCE},\Sigma_j}. \tag{39}$$

Note that the set Σ_1 is empty, as the Hamming distance between two distinct permutations is at least two. The set Σ_2 consists of all transpositions and $|\Sigma_2| = \binom{m}{2} \leq 2^{n(2R)}$. For all $\sigma \in \Sigma_2$, the value of $P_{\mathrm{RCE},\sigma}$ is given by (25), and combining this with (38), we get

$$P_{\mathrm{RCE},\Sigma_2} \leq 2^{-n(R_1(p)-c_n-2R)}. \tag{40}$$

For a given $j>2$, if $\sigma \in \Sigma_j$, then from (36) it follows that $P_{\mathrm{RCE},\sigma} \leq 2^{-nj(\lambda_p-c_n)}$. For $j>2$, the size of the set Σ_j satisfies $|\Sigma_j| < \prod_{i=0}^{j-1}(m-i) < 2^{njR}$. If we define

$$\beta_n \triangleq 2^{-n(\lambda_p-c_n-R)},$$

then we have $P_{\mathrm{RCE},\Sigma_j} \leq \beta_n^j$. Now, if $R<\lambda_p$, then because $c_n = o(1)$, there exists N such that for $n \geq N$, we have $R < \lambda_p - c_n$ and hence $\beta_n < 1$. Therefore, for $n \geq N$,

$$\sum_{j=3}^{m} P_{\mathrm{RCE},\Sigma_j} \leq \sum_{j=3}^{m} \beta_n^j \leq \frac{\beta_n^3}{1-\beta_n}. \tag{41}$$

As $\beta_n \to 0$ and $c_n \to 0$ when $n \to \infty$, it follows from (41) that

$$\sum_{j=3}^{m} P_{\mathrm{RCE},\Sigma_j} \leq \frac{\beta_n^3}{1-\beta_n} \doteq \beta_n^3 \doteq 2^{-3n(\lambda_p-R)}. \tag{42}$$

Combining (39), (40), and (42), for $R < \lambda_p$,

$$\underline{D}(n,R,p) \stackrel{\cdot}{\leq} 2^{-n(R_1(p)-2R)} + 2^{-n(3\lambda_p-3R)}. \quad (43)$$

Comparing (17) with (43), we observe that the error probability $\underline{D}(n,R,p)$ is dominated by $P_{\mathrm{RCE},\sigma}$ terms for σ corresponding to k-cycles with $k=2$ and $k=3$. The next theorem presents an explicit lower bound for $E_{\underline{D}}(R,p)$ when the decoder jointly decodes all the barcodes using a maximum likelihood approach.

Theorem 2: We have

$$E_{\underline{D}}(R,p) \geq |\eta_p(R)|^+, \quad (44)$$

where $\eta_p(R) \triangleq \min\{R_1(p)-2R,\ 2R_0(p)-3R\}$.

Proof: If $R < \lambda_p$, then $R_1(p) \geq 2\lambda_p > 2R$. Therefore, from (43) it follows that if $R < \lambda_p$, then $E_{\underline{D}}(R,p)$ is lower bounded by $\min\{R_1(p)-2R,\ 3\lambda_p-3R\} = \eta_p(R)$. Further, note that $\eta_p(R) > 0$ if and only if $R < \lambda_p$. ∎

The following proposition shows that the lower bound (44) (obtained using joint decoding of barcodes) is *strictly better* than the bound given by (11) (obtained with independent decoding of barcodes) in the interval where it is positive.

Proposition 2: When $R_0(p) > 2R$ *and* $0 < p < 0.5$*, then we have the strict inequality*

$$\eta_p(R) > R_0(p) - 2R.$$

Proof: When $0 < p < 0.5$, we have $0 < 4p(1-p) < \sqrt{4p(1-p)} < 1$, and hence $R_1(p) > R_0(p)$. If $R_0(p) > 2R$, then $2R_0(p) - 3R = 2(R_0(p)-2R) + R > R_0(p) - 2R$. The proof is complete by combining these observations with the definition of $\eta_p(R)$. ∎

Note that $|\eta_p(R)|^+ = 0$ for $R \geq 0.5$, because in this case $\eta_p(R) \leq R_1(p) - 2R \leq R_1(p) - 1 \leq 0$. In the following section, we present improved lower bounds on $E_{\underline{D}}(R,p)$ by analyzing *typical* random codebooks.

III. Typical Random Code

TRCs are known to provide higher error exponents than RCE over a BSC at low rates [8],[11]. Roughly speaking, TRCs are characterized by the property that their relative minimum distance is at least $\delta_{\mathrm{GV}}(2R)$. Formally, for $0 < R < 0.5$, $0 < \epsilon < \delta_{\mathrm{GV}}(2R)$, and indices $1 \leq \hat{\imath} < \hat{\jmath} \leq m = 2^{nR}$, the Hamming distance between codewords $\boldsymbol{c}_{\hat{\imath}}$ and $\boldsymbol{c}_{\hat{\jmath}}$ in a TRC satisfies [8]

$$\Pr\{\mathrm{d_H}(\boldsymbol{c}_{\hat{\imath}},\boldsymbol{c}_{\hat{\jmath}}) = d\} \begin{cases} \leq 2^{-n(1-H(\delta))}, & |\frac{1}{2}-\delta| \leq \frac{1}{2}-\overline{\delta} \\ = 0, & |\frac{1}{2}-\delta| \geq \frac{1}{2}-\underline{\delta}, \end{cases} \quad (45)$$

where $\delta = d/n$, $\overline{\delta} \triangleq \delta_{\mathrm{GV}}(2R)+\epsilon$, and $\underline{\delta} \triangleq \delta_{\mathrm{GV}}(2R)-\epsilon$.

Let $\mathscr{C}_{\mathrm{TRC}}(n,R)$ denote the set of all codebooks of size $2^{nR}\times n$, with the property that the Hamming distance between a pair of codewords $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ satisfies the relation $n\underline{\delta} < \mathrm{d_H}(\boldsymbol{c}_i,\boldsymbol{c}_j) < n(1-\underline{\delta})$ for all $i \neq j$. Note that if codebook C is uniformly distributed over $\mathscr{C}_{\mathrm{TRC}}(n,R)$, then the Hamming distance between a pair of distinct codewords satisfies (45). It is immediate from (4) that

$$\underline{D}(n,R,p) \leq \frac{1}{|\mathscr{C}_{\mathrm{TRC}}(n,R)|} \sum_{C\in\mathscr{C}_{\mathrm{TRC}}(n,R)} D(C,p,\phi), \quad (46)$$

where the expression on the right denotes the average performance using TRCs.

In this section we provide lower bounds on the bee-identification exponent $E_{\underline{D}}(R,p)$ using TRCs. The case where each barcode is decoded independently is analyzed in Sec. III-A while joint barcode decoding is analyzed in Sec. III-B. It is shown that these lower bounds on $E_{\underline{D}}(R,p)$ using TRCs outperform the corresponding bounds for RCEs when the rate is smaller than a certain threshold.

A. Independent Decoding of Barcodes

With independent barcode decoding, the decoder picks $\tilde{\boldsymbol{c}}_j$, the j-th row of $\tilde{C}_\pi$, and then assigns $\nu(j) = \arg\min_k \mathrm{d_H}(\tilde{\boldsymbol{c}}_j,\boldsymbol{c}_k)$, for $1 \leq j \leq m$. From the union bound, we have $D(C,p,\phi) \leq \sum_{j=1}^m \Pr\{\nu(j) \neq \pi^{-1}(j)\}$, and using (46) we get

$$\underline{D}(n,R,p) \leq \sum_{j=1}^{m} \left(\sum_{C\in\mathscr{C}_{\mathrm{TRC}}(n,R)} \frac{\Pr\{\nu(j)\neq\pi^{-1}(j)\}}{|\mathscr{C}_{\mathrm{TRC}}(n,R)|} \right). \quad (47)$$

We now define

$$P_{\mathrm{TRC}}(n,R,p) \triangleq \sum_{C\in\mathscr{C}_{\mathrm{TRC}}(n,R)} \frac{\Pr\{\nu(j)\neq\pi^{-1}(j)\}}{|\mathscr{C}_{\mathrm{TRC}}(n,R)|}.$$

Note that $P_{\mathrm{TRC}}(n,R,p)$ is independent of the index j due to the symmetry resulting from averaging over codebooks uniformly distributed over $\mathscr{C}_{\mathrm{TRC}}(n,R)$. For $i = \pi^{-1}(j)$, the expression for $P_{\mathrm{TRC}}(n,R,p)$ corresponds to the probability of error when the i-th codeword is transmitted. From (47), and the fact that $\underline{D}(n,R,p) \leq 1$, we get

$$\underline{D}(n,R,p) \leq \min\{1,\ mP_{\mathrm{TRC}}(n,R,p)\}. \quad (48)$$

The following theorem uses (48) to present an explicit lower bound on $E_{\underline{D}}(R,p)$ when the rate is smaller than a certain threshold.

Theorem 3: We have

$$E_{\underline{D}}(R,p) \geq \alpha_p \delta_{\mathrm{GV}}(2R), \quad 0 < R < R_{\mathrm{TRC}}(p), \quad (49)$$

where α_p *is defined in* (18)*, and*

$$R_{\mathrm{TRC}}(p) \triangleq 0.5\left(1 - H\left(\frac{\sqrt{4p(1-p)}}{1+\sqrt{4p(1-p)}}\right)\right). \quad (50)$$

Proof: It is known that for $0 < R < R_{\mathrm{TRC}}(p) \leq 0.5$, the error exponent using a TRC over BSC(p), defined as $E_{\mathrm{TRC}}(R,p) \triangleq \liminf_{n\to\infty}(1/n)\log(1/P_{\mathrm{TRC}}(n,R,p))$, is given by [8]

$$E_{\mathrm{TRC}}(R,p) = \alpha_p \delta_{\mathrm{GV}}(2R) + R. \quad (51)$$

Using the fact that $m = 2^{nR}$, and combining (5), (48), with the definition of $E_{\mathrm{TRC}}(R,p)$, we get

$$E_{\underline{D}}(R,p) \geq |E_{\mathrm{TRC}}(R,p) - R|^{+}. \tag{52}$$

The proof is completed by applying (51) in (52). ∎

It is well known that $E_{\mathrm{TRC}}(R,p) > E_{\mathrm{r}}(R,p)$ for $0 < R < R_{\mathrm{TRC}}(p)$ [8]. This implies that the lower bound on $E_{\underline{D}}(R,p)$ for TRC given by (49) is *strictly better* than the corresponding bound for RCE given by (11) when $0 < R < R_{\mathrm{TRC}}(p)$. The next subsection provides a more refined bound on $E_{\underline{D}}(R,p)$ by analyzing joint decoding of barcodes using TRCs.

B. Joint Decoding of Barcodes

With joint barcode decoding, the decoder takes the noisy row-permuted codebook $\tilde{C}_\pi$ as input, and produces the permutation $\nu = \rho^{-1}$ as output, where $\rho = \arg\min_{\sigma \in S_m} \mathrm{d_H}(\tilde{C}_\pi, C_\sigma)$. As in Sec. II-B, we assume, without loss of generality, that the permutation induced by the channel is the identity permutation π_0. For a given codebook C, we have $D(C,p,\phi) \leq \sum_{\sigma \in S_m, \sigma \neq \pi_0} \Pr\{\pi_0 \to \sigma\}$. If we define

$$P_{\mathrm{TRC},\sigma} \triangleq \mathbb{E}\left[\Pr\{\pi_0 \to \sigma\}\right], \tag{53}$$

where the expectation is over a uniform distribution of codebook over $\mathscr{C}_{\mathrm{TRC}}(n,R)$, then we have

$$\begin{aligned} \underline{D}(n,R,p) &\leq \mathbb{E}\left[D(C,p,\phi)\right], \\ &\leq \sum_{\sigma \in S_m, \sigma \neq \pi_0} P_{\mathrm{TRC},\sigma}. \end{aligned} \tag{54}$$

In the following, we quantify $P_{\mathrm{TRC},\sigma}$ for different $\sigma \in S_m$, in order to bound $\underline{D}(n,R,p)$ via (54).

1) σ is a Transposition: If $\sigma = (\hat{\imath}\ \hat{\jmath})$ is the permutation that only transposes indices $\hat{\imath}$ and $\hat{\jmath}$, and $\mathrm{d_H}(\boldsymbol{c}_i, \boldsymbol{c}_j) = d$, then $\mathrm{d_H}\left(C_{\pi_0}, C_{(\hat{\imath}\ \hat{\jmath})}\right) = 2d$, and we have

$$\Pr\{\pi_0 \to (\hat{\imath}\ \hat{\jmath})\} \leq 2^{-2d\alpha_p}. \tag{55}$$

When C is uniformly distributed in $\mathscr{C}_{\mathrm{TRC}}(n,R)$, and $n\underline{\delta} \leq d \leq n(1-\underline{\delta})$, then

$$\begin{aligned} \Pr\left\{\mathrm{d_H}\left(C_{\pi_0}, C_{(\hat{\imath}\ \hat{\jmath})}\right) = 2d\right\} &= \Pr\left\{\mathrm{d_H}(\boldsymbol{c}_{\hat{\imath}}, \boldsymbol{c}_{\hat{\jmath}}) = d\right\}, \\ &\leq 2^{-n(1-H(d/n))}, \end{aligned} \tag{56}$$

where (56) follows from (45). Combining (53), (55), and (56), we get

$$P_{\mathrm{TRC},(\hat{\imath}\ \hat{\jmath})} \leq \sum_{d=n\underline{\delta}}^{n(1-\underline{\delta})} 2^{-n(1-H(d/n)+2(d/n)\alpha_p)}. \tag{57}$$

If $\delta = d/n$ is treated as a continuous variable, then the exponent $E_2(\delta) = 1 - H(\delta) + 2\delta\alpha_p$ is a convex function of δ with a unique minimum at $\hat{\delta}_p$ defined in (22). If we define

$$\hat{R}_p \triangleq 0.5(1 - H(\hat{\delta}_p)), \tag{58}$$

then for $0 < R < \hat{R}_p$, we have

$$\delta_{\mathrm{GV}}(2R) > \delta_{\mathrm{GV}}(2\hat{R}_p) = \hat{\delta}_p.$$

The exponent $E_2(\delta)$ increases monotonically in δ for $\delta \geq \hat{\delta}_p$. Therefore, if $0 < R < \hat{R}_p$ and $\epsilon < \delta_{\mathrm{GV}}(2R) - \hat{\delta}_p$, the exponent in (57) is minimized for $d = n\underline{\delta}$, and we have

$$P_{\mathrm{TRC},(\hat{\imath}\ \hat{\jmath})} \leq 2^{-n(1-H(\underline{\delta})+2\underline{\delta}\alpha_p - c_n)}, \quad 0 < R < \hat{R}_p, \tag{59}$$

where $c_n = (\log(n+1))/n$.

2) σ is a k-Cycle: We now consider the case where σ is a k-cycle with $k \geq 3$. We will apply the following proposition towards characterizing $P_{\mathrm{TRC},\sigma}$.

Proposition 3: Let $\boldsymbol{c}_{i_1}, \boldsymbol{c}_{i_2}, \ldots, \boldsymbol{c}_{i_k}$ be k distinct rows in codebook C, and let d_l satisfy $n\underline{\delta} \leq d_l \leq n(1-\underline{\delta})$ for $1 \leq l \leq k-1$. Let $Q_{\mathrm{TRC}}\left\{\bigcap_{l=1}^{k-1}\left\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\right\}\right\}$ denote the probability $\Pr\left\{\bigcap_{l=1}^{k-1}\left\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\right\}\right\}$ when C is uniformly distributed over $\mathscr{C}_{\mathrm{TRC}}(n,R)$. Then, we have

$$Q_{\mathrm{TRC}}\left\{\bigcap_{l=1}^{k-1}\left\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\right\}\right\} \leq \frac{1}{\alpha_n}\prod_{l=1}^{k-1} 2^{-n(1-H(d_l/n))}, \tag{60}$$

where

$$\alpha_n \triangleq \sum_{(\gamma_1,\gamma_2,\ldots,\gamma_m) \in \mathscr{C}_{\mathrm{TRC}}(n,R)} Q_{\mathrm{RCE}}\left\{\bigcap_{i=1}^{m}\{\boldsymbol{c}_i = \gamma_i\}\right\}, \tag{61}$$

and $Q_{\mathrm{RCE}}\left\{\bigcap_{i=1}^{m}\{\boldsymbol{c}_i = \gamma_i\}\right\}$ denotes the probability $\Pr\left\{\bigcap_{i=1}^{m}\{\boldsymbol{c}_i = \gamma_i\}\right\}$ when C is uniformly distributed over $\mathscr{C}(n,R)$.

Proof: See Appendix B. ∎

Now, given that $\sigma = (i_1\ i_2\ \cdots\ i_k)$ and $\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l$ for $1 \leq l \leq k-1$, and $\mathrm{d_H}(\boldsymbol{c}_{i_k}, \boldsymbol{c}_{i_1}) = d_k$, we have $\mathrm{d_H}(C_{\pi_0}, C_\sigma) = \sum_{l=1}^{k} d_l$, and therefore

$$\Pr\{\pi_0 \to \sigma\} \leq 2^{-\left(\sum_{l=1}^{k} d_l\right)\alpha_p}. \tag{62}$$

If $d_0 \triangleq n\underline{\delta}$, then combining (60) and (62), we get

$$\begin{aligned} P_{\mathrm{TRC},\sigma} &\leq \sum_{\substack{d_0 \leq d_l \leq n-d_0, \\ 1 \leq l \leq k}} \left(2^{-n\left(\sum_{l=1}^{k}(d_l/n)\alpha_p\right)} \right. \\ &\qquad \left. \times \frac{1}{\alpha_n} 2^{-n\left(\sum_{l=1}^{k-1}(1-H(d_l/n))\right)}\right), \\ &= \frac{1}{\alpha_n}\,\eta_k \prod_{l=1}^{k-1}\zeta_l, \end{aligned} \tag{63}$$

where, for $1 \leq l \leq k-1$, we have

$$\zeta_l \triangleq \sum_{d_0 \leq d_l \leq n-d_0} 2^{-n(1-H(d_l/n)+(d_l/n)\alpha_p)}, \tag{64}$$

and

$$\eta_k \triangleq \sum_{d_0 \leq d_k \leq n-d_0} 2^{-d_k\alpha_p} \leq 2^{-n(\underline{\delta}\alpha_p - c_n)}. \tag{65}$$

The function $E_1(\delta) = 1 - H(\delta) + \delta\alpha_p$ is a convex function of δ, and has a unique minimum that occurs at $\tilde{\delta}_p$ defined in (32). From (50) we observe that $R_{\mathrm{TRC}}(p) = 0.5(1 - H(\tilde{\delta}_p))$. Thus, if $R < R_{\mathrm{TRC}}(p)$, then we have $\delta_{\mathrm{GV}}(2R) > \tilde{\delta}_p$. Further, $E_1(\delta)$ is an increasing function of δ for $\delta \geq \tilde{\delta}_p$, and so if

$R < R_{\mathrm{TRC}}(p)$ and $\epsilon < \delta_{\mathrm{GV}}(2R) - \tilde{\delta}_p$, the exponent in (64) is minimized when $d_l = d_0 = n\underline{\delta}$. Thus, we have

$$\zeta_l \leq 2^{-n(1-H(\underline{\delta})+\underline{\delta}\alpha_p - c_n)}, \quad 0 < R < R_{\mathrm{TRC}}(p). \tag{66}$$

Combining (63), (65), and (66), for $0 < R < R_{\mathrm{TRC}}(p)$,

$$P_{\mathrm{TRC},\sigma} \leq \frac{1}{\alpha_n} 2^{-n((k-1)(1-H(\underline{\delta}))+k(\underline{\delta}\alpha_p - c_n))}, \tag{67}$$

where σ is a k-cycle with $k > 2$. As $k < 2(k-1)$ for $k > 2$, it follows from (67) that

$$P_{\mathrm{TRC},\sigma} \leq \frac{1}{\alpha_n} 2^{-nk(0.5(1-H(\underline{\delta}))+\underline{\delta}\alpha_p - c_n)}, \quad 0 < R < R_{\mathrm{TRC}}(p). \tag{68}$$

Recall that $\hat{\delta}_p$ and $\hat{R}_p$ are given by (22) and (58), respectively. As $x/(1+x)$ is an increasing function of x, and $0 < p < 0.5$, it follows that $\hat{\delta}_p < \tilde{\delta}_p < 0.5$, which implies that $R_{\mathrm{TRC}}(p) < \hat{R}_p$. Note that a transposition is simply a k-cycle with $k = 2$, and comparing (59) with (68) we observe that the relation given by (68) holds even when $k = 2$.

3) σ is a Product (Composition) of Two Disjoint Cycles: We now consider the case where $\sigma = \sigma_1\sigma_2$, where σ_1 and σ_2 are disjoint cycles of length k_1 and k_2, respectively. Let $\sigma_1 = (i_1\ i_2\ \cdots\ i_{k_1})$ and $\sigma_2 = (i_{k_1+1}\ i_{k_1+1}\ \cdots\ i_{k_1+k_2})$. If $d_0 \leq d_l \leq n - d_0$ for $1 \leq l \leq k_1 + k_2$, then a straightforward extension of Prop. 3 shows that the probability

$$\Pr\Big\{ \bigcap_{l=1}^{k_1-1} \{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\} \bigcap \{\mathrm{d_H}(\boldsymbol{c}_{i_{k_1}}, \boldsymbol{c}_{i_1}) = d_{k_1}\} \bigcap_{l=k_1+1}^{k_1+k_2-1} \{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\} \bigcap \{\mathrm{d_H}(\boldsymbol{c}_{i_{k_1+k_2}}, \boldsymbol{c}_{i_{k_1+1}}) = d_{k_1+k_2}\} \Big\}$$

is upper bounded by

$$\frac{1}{\alpha_n} 2^{-n\left(\sum_{l=1}^{k_1-1}(1-H(d_l/n))\right)} \times 2^{-n\left(\sum_{l=k_1+1}^{k_1+k_2-1}(1-H(d_l/n))\right)}. \tag{69}$$

Further, for a given codebook C, with $\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l$, $1 \leq l \leq k_1 - 1$, $\mathrm{d_H}(\boldsymbol{c}_{i_{k_1}}, \boldsymbol{c}_{i_1}) = d_{k_1}$, $\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l$, $k_1 + 1 \leq l \leq k_1 + k_2 - 1$, $\mathrm{d_H}(\boldsymbol{c}_{i_{k_1+k_2}}, \boldsymbol{c}_{i_{k_1+1}}) = d_{k_1+k_2}$, we have $\mathrm{d_H}(C_{\pi_0}, C_\sigma) = \sum_{l=1}^{k_1+k_2} d_l$, and therefore

$$\Pr\{\pi_0 \to \sigma\} \leq 2^{-\left(\sum_{l=1}^{k_1+k_2} d_l\right)\alpha_p}. \tag{70}$$

Combining (69) and (70), we can upper bound $P_{\mathrm{TRC},\sigma}$ by

$$\frac{1}{\alpha_n} \sum_{\substack{d_0 \leq d_l \leq n-d_0,\\ 1 \leq l \leq k_1+k_2}} \left(2^{-n\left((\sum_{l=1}^{k_1+k_2}(d_l/n)\alpha_p)\right)} \times 2^{-n\left(\sum_{l=1}^{k_1-1}(1-H(d_l/n)) + \sum_{l=k_1+1}^{k_1+k_2-1}(1-H(d_l/n))\right)} \right). \tag{71}$$

The above expression can be equivalently written as

$$\frac{1}{\alpha_n} {\eta_k}^2 {\zeta_l}^{k_1+k_2-2}, \tag{72}$$

where ζ_l and η_k are defined in (64) and (65), respectively. Now, applying (65), (66) in (72) for $0 < R < R_{\mathrm{TRC}}(p)$, we get

$$P_{\mathrm{TRC},\sigma} \leq \frac{1}{\alpha_n} 2^{-n((k_1+k_2-2)(1-H(\underline{\delta}))+(k_1+k_2)(\underline{\delta}\alpha_p - c_n))}, \tag{73}$$

where $\sigma = (i_1\ i_2\ \cdots\ i_{k_1})(i_{k_1+1}\ i_{k_1+2}\ \cdots\ i_{k_1+k_2})$. As $k_1 \geq 2$ and $k_2 \geq 2$, we have $2(k_1 + k_2 - 2) \geq k_1 + k_2$, and therefore for $0 < R < R_{\mathrm{TRC}}(p)$, we have

$$P_{\mathrm{TRC},\sigma} \leq \frac{1}{\alpha_n} 2^{-n(k_1+k_2)(0.5(1-H(\underline{\delta}))+\underline{\delta}\alpha_p - c_n)}. \tag{74}$$

4) General $\sigma \in S_m$ With $\sigma \neq \pi_0$: If permutation σ is a product of r disjoint cycles of length $k_1, \ldots, k_r$, respectively, then similar to (68), (74), we have for $0 < R \leq R_{\mathrm{TRC}}(p)$,

$$P_{\mathrm{TRC},\sigma} \leq \frac{1}{\alpha_n} 2^{-n\left(\sum_{t=1}^{r} k_t\right)(0.5(1-H(\underline{\delta}))+\underline{\delta}\alpha_p - c_n)}. \tag{75}$$

5) Putting it all Together: For $1 \leq j \leq m$, if we define $P_{\mathrm{TRC},\Sigma_j} \triangleq \sum_{\sigma\in\Sigma_j} P_{\mathrm{TRC},\sigma}$, where Σ_j is given by (37), then (54) can be equivalently expressed as

$$\underline{D}(n, R, p) \leq \sum_{j=2}^{m} P_{\mathrm{TRC},\Sigma_j}. \tag{76}$$

If σ is a product of r disjoint cycles of length $k_1, \ldots, k_r$, respectively, and $s = \sum_{t=1}^{r} k_t$, then σ belongs to the set Σ_s, and $P_{\mathrm{TRC},\sigma}$ is given by (75). Equivalently, for a given $j \geq 2$, if $\sigma \in S_m$ belongs to the set Σ_j, then for $0 < R < R_{\mathrm{TRC}}(p)$,

$$P_{\mathrm{TRC},\sigma} \leq \frac{1}{\alpha_n} 2^{-nj(0.5(1-H(\underline{\delta}))+\underline{\delta}\alpha_p - c_n)}. \tag{77}$$

The size of Σ_j satisfies $|\Sigma_j| < \prod_{i=0}^{j-1}(m-i) < 2^{njR}$. Therefore, for $0 < R < R_{\mathrm{TRC}}(p)$, we have

$$\begin{aligned} P_{\mathrm{TRC},\Sigma_j} &= \sum_{\sigma\in\Sigma_j} P_{\mathrm{TRC},\sigma} \\ &\leq \frac{1}{\alpha_n} 2^{-nj(0.5(1-H(\underline{\delta}))+\underline{\delta}\alpha_p - c_n)}\, 2^{njR} \\ &= \frac{1}{\alpha_n} 2^{-nj(0.5(1-H(\underline{\delta}))-R+\underline{\delta}\alpha_p - c_n)}. \end{aligned} \tag{78}$$

Now, if we define $\xi_n \triangleq 2^{-n(0.5(1-H(\underline{\delta}))-R+\underline{\delta}\alpha_p - c_n)}$, then (78) can be equivalently expressed as $P_{\mathrm{TRC},\Sigma_j} \leq (1/\alpha_n)\xi_n^j$. As $c_n = o(1)$, there exists $\hat{N}$ such that for $n \geq \hat{N}$, we have $c_n < 0.5(1 - H(\underline{\delta})) - R + \underline{\delta}\alpha_p$ and hence $\xi_n < 1$. Therefore, for $n \geq \hat{N}$ and $0 < R < R_{\mathrm{TRC}}(p)$, we have

$$\begin{aligned} \underline{D}(n, R, p) &\leq \frac{1}{\alpha_n} \sum_{j=2}^{m} \xi_n^j \\ &< \frac{1}{\alpha_n} \frac{\xi_n^2}{1-\xi_n} \\ &\doteq \frac{\xi_n^2}{1-\xi_n} && (79) \\ &\doteq \xi_n^2 && (80) \\ &= 2^{-n(1-H(\underline{\delta})-2R+2\underline{\delta}\alpha_p - 2c_n)} \\ &\doteq 2^{-n(1-H(\underline{\delta})-2R+2\underline{\delta}\alpha_p)}, && (81) \end{aligned}$$

where (79) follows because $\alpha_n \to 1$ as $n \to \infty$ [8], and (80) follows because $\xi_n = o(1)$, while (81) follows because

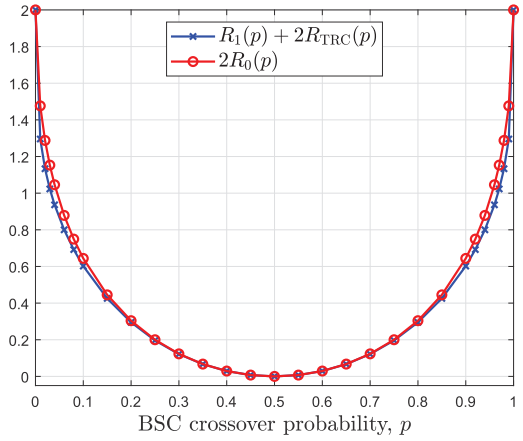

Fig. 4. Plot demonstrating $2R_0(p) \geq R_1(p) + 2R_{\mathrm{TRC}}(p)$.

$c_n = o(1)$. Note that $\underline{\delta} = \delta_{\mathrm{GV}}(2R) - \epsilon$, and so $\lim_{\epsilon \to 0} \underline{\delta} = \delta_{\mathrm{GV}}(2R)$ and $\lim_{\epsilon \to 0} (1 - H(\underline{\delta}) - 2R + 2\underline{\delta}\alpha_p) = 2\delta_{\mathrm{GV}}(2R)\alpha_p$. As ϵ can be made arbitrarily small, it follows from (81) that for $0 < R < R_{\mathrm{TRC}}(p)$, we have

$$\underline{D}(n, R, p) \stackrel{\cdot}{\leq} 2^{-n(2\delta_{\mathrm{GV}}(2R)\alpha_p)}. \tag{82}$$

The following theorem encapsulates the main result of this subsection on bounding the bee-identification exponent, $E_{\underline{D}}(R, p)$, using joint decoding for TRC.

Theorem 4: We have

$$E_{\underline{D}}(R, p) \geq 2\delta_{\mathrm{GV}}(2R)\,\alpha_p, \quad 0 < R < R_{\mathrm{TRC}}(p). \tag{83}$$

Proof: Follows from (5) and (82). ∎

We note that the above lower bound for $E_{\underline{D}}(R, p)$ using TRCs with joint barcode decoding is *twice* the corresponding bound obtained using independent barcode decoding (see (49)). The following proposition shows that the lower bound given by Thm. 4 using TRC is *strictly better* than corresponding bound using RCE (see Thm. 2) for $0 < R < R_{\mathrm{TRC}}(p)$.

Proposition 4: The lower bound on $E_{\underline{D}}(R, p)$ in (83) *obtained for TRC is strictly better than the corresponding bound in* (44) *obtained for RCE when* $0 < R < R_{\mathrm{TRC}}(p)$.

Proof: It is known that $E_{\mathrm{TRC}}(R, p) > E_{\mathrm{r}}(R, p)$ when $0 < R < R_{\mathrm{TRC}}(p)$ [8]. Further, using explicit numerical computation, it can be shown that $2R_0(p) \geq R_1(p) + 2\ R_{\mathrm{TRC}}(p)$ (see Fig. 4). Therefore, it follows that for $0 < R < R_{\mathrm{TRC}}(p)$, we have

$$\begin{aligned} 2\delta_{\mathrm{GV}}(2R)\,\alpha_p &= 2\,(E_{\mathrm{TRC}}(R, p) - R) \\ &> 2\,(E_{\mathrm{r}}(R, p) - R) = 2(R_0(p) - 2R) \\ &\geq R_1(p) - 2R + 2(R_{\mathrm{TRC}}(p) - R) \\ &> R_1(p) - 2R \geq \eta_p(R). \end{aligned}$$ ∎

The next section presents an explicit upper bound for $E_{\underline{D}}(R, p)$ which applies to all possible codebook designs.

IV. Upper Bound on the Bee-Identification Exponent

This section presents an upper bound on the bee-identification exponent $E_{\underline{D}}(R, p)$. Towards this, we define the following optimum minimum distance metrics

$$\begin{aligned} d^*(n, R) &\triangleq \max_{C \in \mathscr{C}(n,R)} \min_{\substack{\boldsymbol{c}_i, \boldsymbol{c}_j \in C \\ \boldsymbol{c}_i \neq \boldsymbol{c}_j}} \mathrm{d_H}(\boldsymbol{c}_i, \boldsymbol{c}_j), \\ \delta^*(n, R) &\triangleq d^*(n, R)/n, \\ \delta^*(R) &\triangleq \limsup_{n \to \infty} \delta^*(n, R). \end{aligned}$$

The upper bound on the bee-identification exponent, given by Theorem 5, relies on the existence of a set consisting of at least $m/4$ disjoint pairs of codeword indices (where m is the total number of codewords in the codebook), such that for every pair of indices, the corresponding codewords have sufficiently small Hamming distance. In particular, for any given codebook $C \in \mathscr{C}(n, R)$, we show that there exists a set $\mathscr{I}_C$ consisting of pairs of codeword indices $(i, j), i \neq j$, satisfying the following properties:

(i) If $(i, j) \in \mathscr{I}_C$, then $\mathrm{d_H}(\boldsymbol{c}_i, \boldsymbol{c}_j) \leq d^*(n, R - \frac{1}{n})$.
(ii) If $(i, j) \in \mathscr{I}_C$ and $(\hat{\imath}, \hat{\jmath}) \in \mathscr{I}_C$, then $\hat{\imath} \neq i, \hat{\imath} \neq j$ and $\hat{\jmath} \neq i, \hat{\jmath} \neq j$.
(iii) Size of set $\mathscr{I}_C$ is at least $m/4$.

A set satisfying the above properties can be constructed iteratively as follows.

- *Step 1*: For a given codebook $C \in \mathscr{C}(n, R)$, initialize $\mathscr{I}_C$ to be the empty set and let $\mathcal{T} = C$.
- *Step 2*: As $\mathcal{T}$ contains at least $m/2 = 2^{n(R - \frac{1}{n})}$ codewords, it follows from the definition of $d^*(n, R - \frac{1}{n})$ that there exists distinct $\boldsymbol{c}_i, \boldsymbol{c}_j \in \mathcal{T}$, satisfying $\mathrm{d_H}(\boldsymbol{c}_i, \boldsymbol{c}_j) \leq d^*(n, R - \frac{1}{n})$. Include the pair (i, j) to $\mathscr{I}_C$, and let $\mathcal{T} = \mathcal{T} \setminus \{\boldsymbol{c}_i, \boldsymbol{c}_j\}$.
- *Step 3*: If $|\mathscr{I}_C| < m/4$, then go to *Step 2*, else stop.

Let the receiver employ ML decoding, and interpret each pair $(i, j) \in \mathscr{I}_C$ as a transposition $\sigma = (i\ j)$ that interchanges indices i and j. Let $A_{(i,j)}$ denote the error event that the receiver incorrectly decodes the channel induced permutation to transposition $(i\ j)$ (instead of the identity permutation π_0), i.e. $A_{(i,j)} = \{\pi_0 \to (i\ j)\}$. Then, the bee-identification error probability $D(C, p, \phi)$ can be lower bounded as

$$D(C, p, \phi) \geq \Pr\left\{ \bigcup_{(i,j) \in \mathscr{I}_C} A_{(i,j)} \right\}. \tag{84}$$

Using de Caen's lower bound on the probability of a union [12], the expression on the right side in (84) can itself be lower bounded by

$$\begin{aligned} &\sum_{(i,j) \in \mathscr{I}_C} \frac{\left(\Pr\{A_{(i,j)}\}\right)^2}{\Pr\{A_{(i,j)}\} + \sum\limits_{\substack{(\hat{\imath},\hat{\jmath}) \in \mathscr{I}_C \\ (\hat{\imath},\hat{\jmath}) \neq (i,j)}} \Pr\left\{A_{(i,j)} \cap A_{(\hat{\imath},\hat{\jmath})}\right\}}, \\ &\stackrel{(a)}{=} \sum_{(i,j) \in \mathscr{I}_C} \frac{\left(\Pr\{A_{(i,j)}\}\right)^2}{\Pr\{A_{(i,j)}\} + \sum\limits_{\substack{(\hat{\imath},\hat{\jmath}) \in \mathscr{I}_C \\ (\hat{\imath},\hat{\jmath}) \neq (i,j)}} \Pr\left\{A_{(i,j)}\right\} \Pr\left\{A_{(\hat{\imath},\hat{\jmath})}\right\}}, \\ &\geq \frac{\sum\limits_{(i,j) \in \mathscr{I}_C} \Pr\{A_{(i,j)}\}}{1 + \sum\limits_{(\hat{\imath},\hat{\jmath}) \in \mathscr{I}_C} \Pr\left\{A_{(\hat{\imath},\hat{\jmath})}\right\}}, \end{aligned} \tag{85}$$

where (a) follows because events $A_{(i,j)}$ and $A_{(\hat{\imath},\hat{\jmath})}$ are independent when the sets $\{i,j\}$ and $\{\hat{\imath},\hat{\jmath}\}$ are disjoint. Now

$$\begin{aligned}\sum_{(i,j)\in\mathscr{I}_C}\Pr\{A_{(i,j)}\} &\overset{(b)}{\dot{\geq}} \sum_{(i,j)\in\mathscr{I}_C} 2^{-n\left(2\delta^*\left(n,R-\frac{1}{n}\right)\alpha_p\right)},\\ &\doteq \sum_{(i,j)\in\mathscr{I}_C} 2^{-n(2\delta^*(n,R)\alpha_p)},\\ &\overset{(c)}{\geq} 2^{-n\left(2\delta^*(n,R)\alpha_p-\left(R-\frac{2}{n}\right)\right)},\\ &\doteq 2^{-n(2\delta^*(R)\alpha_p-R)}, \end{aligned} \tag{86}$$

where (b) follows from the fact that $\mathrm{d_H}(C_{\pi_0},C_{(i,j)}) \leq 2\,d^*(n,R-\frac{1}{n})$ for $(i,j)\in\mathscr{I}_C$, and (c) follows because $|\mathscr{I}_C|\geq m/4$. If $R_{\mathrm{UB}}(p)\triangleq\sup\{R: 2\delta^*(R)\alpha_p > R\}$, then combining (84), (85), (86), and noting that $x/(1+x)$ increases with x, we have

$$\begin{aligned}D(C,p,\phi) &\dot{\geq} \frac{2^{-n(2\delta^*(R)\alpha_p-R)}}{1+2^{-n(2\delta^*(R)\alpha_p-R)}},\\ &\doteq 2^{-n(2\delta^*(R)\alpha_p-R)}, \quad 0<R<R_{\mathrm{UB}}(p). \end{aligned} \tag{87}$$

As (87) is true for all $C\in\mathscr{C}(n,R)$, we have

$$\underline{D}(n,R,p)\dot{\geq} 2^{-n(2\delta^*(R)\alpha_p-R)}, \quad 0<R<R_{\mathrm{UB}}(p). \tag{88}$$

The value $\delta^*(R)$ can be upper bounded as [13], [14]

$$\delta^*(R)\leq\delta_{\mathrm{LP}}(R)\triangleq\frac{1}{2}-\sqrt{\delta_{\mathrm{GV}}(1-R)(1-\delta_{\mathrm{GV}}(1-R))}. \tag{89}$$

The following theorem provides an upper bound on the bee-identification exponent $E_{\underline{D}}(R,p)$.

Theorem 5: We have

$$E_{\underline{D}}(R,p)\leq|2\delta^*(R)\alpha_p-R|^+\leq|2\delta_{\mathrm{LP}}(R)\alpha_p-R|^+. \tag{90}$$

Proof: Follows immediately from (88) and (89). ∎

The following corollary shows that $E_{\underline{D}}(R,p)$ can be explicitly characterized with a rather simple expression when rate R tends to zero.

Corollary 1: We have

$$\lim_{R\to0}E_{\underline{D}}(R,p)=\alpha_p. \tag{91}$$

Proof: As $\lim_{R\to0}\delta_{\mathrm{LP}}(R)=0.5$, we have from (90) that

$$\lim_{R\to0}E_{\underline{D}}(R,p)\leq\lim_{R\to0}(2\delta_{\mathrm{LP}}(R)\alpha_p-R)=\alpha_p. \tag{92}$$

On the other hand, we have $\lim_{R\to0}\delta_{\mathrm{GV}}(R)=0.5$ and so it follows from (83) that

$$\lim_{R\to0}E_{\underline{D}}(R,p)\geq\lim_{R\to0}2\delta_{\mathrm{GV}}(2R)\alpha_p=\alpha_p. \tag{93}$$

The proof is completed by using (92) and (93). ∎

The above corollary shows that the lower bound on $E_{\underline{D}}(R,p)$ given by (83), and the upper bound on $E_{\underline{D}}(R,p)$ given by (90) become *tight* as $R\to0$.

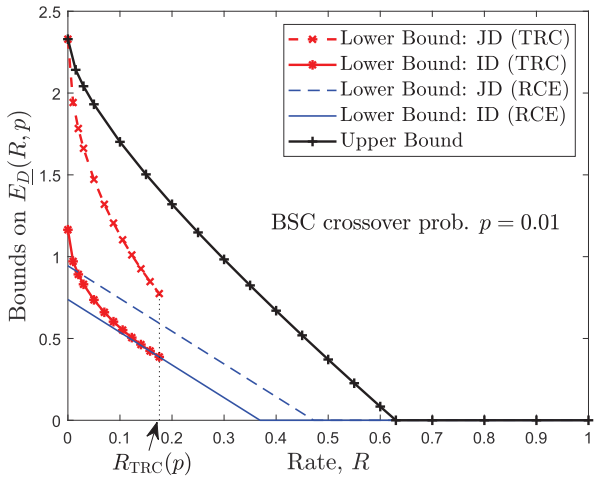

Fig. 5. Lower bounds on $E_{\underline{D}}(R,p)$ with independent decoding (ID) and joint decoding (JD) using TRC and RCE. The upper bound holds for all codebook designs.

V. A Numerical Example

Fig. 5 plots different bounds for the bee-identification exponent $E_{\underline{D}}(R,p)$. The explicit lower bound for RCE with independent decoding (ID) (respectively, joint decoding (JD)) is given by (11) (respectively, (44)). The performance with JD is seen to be much better than with ID. When $0<R<R_{\mathrm{TRC}}(p)$, the explicit lower bound for TRC with ID (respectively, JD) is given by (49) (respectively, (83)). As shown in Prop. 4, the lower bound obtained using TRC with joint decoding is better than the corresponding bound using RCE. The upper bound is given by (90) and holds for all possible codebook designs. Further, as shown in Cor. 1, it is observed from Fig. 5 that $\lim_{R\to0}E_{\underline{D}}(R,p)=\alpha_p=2.33$ for $p=0.01$.

VI. Discussion

We introduced the information-theoretic "bee-identification problem" which arises naturally in different massive identification settings. We derived explicit upper and lower bounds on the bee-identification exponent, and showed that joint decoding of barcodes provides a significantly better exponent than separate decoding followed by permutation inference. For low rates, we showed that the lower bound on the bee-identification exponent obtained using TRC is strictly better than the corresponding bound obtained using RCE. Moreover, when the rate approaches zero, we showed that the upper bound on the bee-identification exponent coincides with the lower bound obtained using TRC with joint barcode decoding.

Relative to the independent decoding of barcodes, the performance improvement with joint decoding comes at a cost of increased computational complexity. For joint decoding, an exhaustive search entails comparing the received noisy & permuted version of the codebook with $m!$ row-permutations of the codebook. This may be computationally prohibitive even for moderate values of blocklength n when m scales exponentially with n. In practice, intermediate performance between the extremes of independent decoding and joint decoding may be achieved with manageable complexity using ideas from generalized minimum distance decoding [15]. In particular,

the decoding process may proceed in two steps: The first step involves independent decoding of each barcode where an erasure is declared if the distance between the received noisy barcode to the nearest barcode in the codebook exceeds a threshold. The second step fixes the codebook row-indices corresponding to the un-erased barcodes, and then decodes the erased barcodes by jointly comparing their received noisy version to different row-permutations of the codebook corresponding to the non-fixed indices. This results in significant reduction in complexity in case only a few barcodes are declared as erasure in the first step. Therefore, we have a tradeoff between performance and complexity via an appropriate choice of the distance threshold parameter for declaring an erasure.

The work in this paper may be extended by considering different variants of the bee-identification error metric, for instance, where error is flagged only when the fraction of incorrectly decoded barcodes exceeds a threshold. Another interesting scenario for future analysis is the problem formulation where some of the m rows in codebook C are deleted, due to some bees being outside the hive when taking the picture.

APPENDIX A
PROOF OF PROP. 1

Proof: Let $\mathbb{F}_{2^n}$ denote the space of all n-length binary vectors, and let $\gamma_{k-1}, \tilde{\gamma}_{k-1} \in \mathbb{F}_{2^n}$, and $\Delta \triangleq \gamma_{k-1} \oplus \tilde{\gamma}_{k-1}$, where $\oplus$ denotes modulo-2 addition. Note that when codebook C is uniformly distributed over $\mathscr{C}(n,R)$, then the rows $\boldsymbol{c}_{i_l}$, for $1 \leq l \leq k$, are i.i.d. and uniformly distributed over $\mathbb{F}_{2^n}$. We have $\Pr\left\{\mathrm{d_H}(\gamma_{k-1}, \boldsymbol{c}_{i_k}) = d_{k-1}\right\} = \Pr\left\{\mathrm{d_H}(\tilde{\gamma}_{k-1}, \boldsymbol{c}_{i_k} + \Delta) = d_{k-1}\right\} \overset{\text{(i)}}{=} \Pr\left\{\mathrm{d_H}(\tilde{\gamma}_{k-1}, \boldsymbol{c}_{i_k}) = d_{k-1}\right\}$, where (i) follows from the fact that for a given Δ, the distribution of $\boldsymbol{c}_{i_k}+\Delta$ is same as the distribution of $\boldsymbol{c}_{i_k}$. This implies that $\Pr\left\{\mathrm{d_H}(\boldsymbol{c}_{i_{k-1}}, \boldsymbol{c}_{i_k}) = d_{k-1} | \boldsymbol{c}_{i_{k-1}} = \gamma_{k-1}\right\} \overset{\text{(ii)}}{=} \Pr\left\{\mathrm{d_H}(\boldsymbol{c}_{i_{k-1}}, \boldsymbol{c}_{i_k}) = d_{k-1}\right\}$. Then $\Pr\left\{\bigcap_{l=1}^{k-1}\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\}\right\}$ can be expressed as

$$\begin{aligned}
&\sum_{\gamma_1,\ldots,\gamma_{k-1}\in\mathbb{F}_{2^n}} \left(\Pr\left\{\bigcap_{l=1}^{k-1}\{\boldsymbol{c}_{i_l} = \gamma_l\}\right\} \right. \\
&\quad \left. \times \Pr\left\{\bigcap_{l=1}^{k-1}\left\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\right\} \,\middle|\, \bigcap_{l=1}^{k-1}\{\boldsymbol{c}_{i_l} = \gamma_l\}\right\} \right), \\
&= \sum_{\gamma_1,\ldots,\gamma_{k-1}} \left(\Pr\left\{\bigcap_{l=1}^{k-1}\{\boldsymbol{c}_{i_l} = \gamma_i\}\right\} \mathbf{1}_{\left\{\bigcap_{l=1}^{k-2}\{\mathrm{d_H}(\gamma_l,\gamma_{l+1})=d_l\}\right\}} \right. \\
&\quad \left. \times \Pr\left\{\mathrm{d_H}(\boldsymbol{c}_{i_{k-1}}, \boldsymbol{c}_{i_k}) = d_{k-1} \middle| \boldsymbol{c}_{i_{k-1}} = \gamma_{k-1}\right\} \right), \\
&\overset{\text{(iii)}}{=} \sum_{\gamma_1,\ldots,\gamma_{k-1}} \left(\Pr\left\{\bigcap_{l=1}^{k-1}\{\boldsymbol{c}_{i_l} = \gamma_l\}\right\} \mathbf{1}_{\left\{\bigcap_{l=1}^{k-2}\{\mathrm{d_H}(\gamma_l,\gamma_{l+1})=d_l\}\right\}} \right. \\
&\quad \left. \times \Pr\left\{\mathrm{d_H}(\boldsymbol{c}_{i_{k-1}}, \boldsymbol{c}_{i_k}) = d_{k-1}\right\} \right), \\
&= \Pr\left\{\bigcap_{l=1}^{k-2} \mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\right\} \Pr\left\{\mathrm{d_H}(\boldsymbol{c}_{i_{k-1}}, \boldsymbol{c}_{i_k}) = d_{k-1}\right\},
\end{aligned} \tag{94}$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function, and (iii) follows from (ii). Recursively applying (94), we get

$$\Pr\left\{\bigcap_{l=1}^{k-1}\left\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\right\}\right\} = \prod_{l=1}^{k-1} \Pr\left\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\right\}.$$

Now, (28) follows from the fact that $\Pr\left\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\right\} \leq 2^{-n(1-H(d_l/n))}$ when $\boldsymbol{c}_{i_l}$ and $\boldsymbol{c}_{i_{l+1}}$ are uniformly distributed over $\mathbb{F}_{2^n}$ [8]. ■

APPENDIX B
PROOF OF PROP. 3

Proof: For $1 \leq i \leq m = 2^{nR}$, let $\boldsymbol{c}_i$ denote the i-th row of codebook C. Let $\mathbb{F}_{2^n}$ denote the space of all n-length binary vectors, and let $\gamma_i \in \mathbb{F}_{2^n}$ for $1 \leq i \leq m$. Let $Q_{\mathrm{TRC}}\{\bigcap_{i=1}^m\{\boldsymbol{c}_i = \gamma_i\}\}$ denote the probability $\Pr\{\bigcap_{i=1}^m\{\boldsymbol{c}_i = \gamma_i\}\}$ when C is uniformly distributed over $\mathscr{C}_{\mathrm{TRC}}(n,R)$. Then, we have

$$\begin{aligned}
&Q_{\mathrm{TRC}}\left\{\bigcap_{i=1}^m\{\boldsymbol{c}_i = \gamma_i\}\right\} \\
&\quad = \frac{1}{\alpha_n} Q_{\mathrm{RCE}}\left\{\bigcap_{i=1}^m\{\boldsymbol{c}_i = \gamma_i\}\right\} \mathbf{1}_{\{(\gamma_1,\gamma_2,\ldots,\gamma_m)\in\mathscr{C}_{\mathrm{TRC}}(n,R)\}},
\end{aligned} \tag{95}$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function. Further, let $Q_{\mathrm{RCE}}\left\{\bigcap_{l=1}^{k-1}\left\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\right\}\right\}$ denote the probability $\Pr\left\{\bigcap_{l=1}^{k-1}\left\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\right\}\right\}$ when codebook C is uniformly distributed over $\mathscr{C}(n,R)$. Then,

$$\begin{aligned}
&Q_{\mathrm{TRC}}\left\{\bigcap_{l=1}^{k-1}\left\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\right\}\right\} \\
&= \sum_{\substack{\gamma_i\in\mathbb{F}_{2^n},\\ 1\leq i\leq m}} Q_{\mathrm{TRC}}\left\{\bigcap_{i=1}^m\{\boldsymbol{c}_i = \gamma_i\}\right\} \mathbf{1}_{\left\{\bigcap_{l=1}^{k-1}\mathrm{d_H}(\gamma_{i_l},\gamma_{i_{l+1}})=d_l\right\}}, \\
&\overset{(a)}{\leq} \frac{1}{\alpha_n} \sum_{\substack{\gamma_i\in\mathbb{F}_{2^n},\\ 1\leq i\leq m}} Q_{\mathrm{RCE}}\left\{\bigcap_{i=1}^m\{\boldsymbol{c}_i = \gamma_i\}\right\} \mathbf{1}_{\left\{\bigcap_{l=1}^{k-1}\mathrm{d_H}(\gamma_{i_l},\gamma_{i_{l+1}})=d_l\right\}}, \\
&= \frac{1}{\alpha_n} Q_{\mathrm{RCE}}\left\{\bigcap_{l=1}^{k-1}\left\{\mathrm{d_H}(\boldsymbol{c}_{i_l}, \boldsymbol{c}_{i_{l+1}}) = d_l\right\}\right\}, \\
&\overset{(b)}{\leq} \frac{1}{\alpha_n} \prod_{l=1}^{k-1} 2^{-n(1-H(d_l/n))},
\end{aligned}$$

where (a) follows from (95), and (b) follows from Prop. 1. ■

ACKNOWLEDGMENT

The authors acknowledge discussions with Ting-Yi Wu, Tim Gernat, and Prof. Gene Robinson on the bee-identification problem formulation. The authors would also like to thank Prof. Neri Merhav (Technion) for suggesting several improvements to the manuscript and, in particular, to refine the asymptotic statements herein.

REFERENCES

[1] T. Gernat, V. D. Rao, M. Middendorf, H. Dankowicz, N. Goldenfeld, and G. E. Robinson, "Automated monitoring of behavior reveals bursty interaction patterns and rapid spreading dynamics in honeybee social networks," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 7, pp. 1433–1438, Feb. 2018.

[2] S. Shahi, D. Tuninetti, and N. Devroye, "The strongly asynchronous massive access channel," Jul. 2018, *arXiv:1807.09934*. [Online]. Available: https://arxiv.org/abs/1807.09934

[3] S. Shahi, D. Tuninetti, and N. Devroye, "On identifying a massive number of distributions," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 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. Inf. 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. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 3130–3134.

[6] I. Shomorony and R. Heckel, "Capacity results for the noisy shuffling channel," in *Proc. 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. Inf. Theory*, vol. 64, no. 7, pp. 5156–5169, Jul. 2018.

[8] A. Barg and G. D. Forney, Jr., "Random codes: Minimum distances and error exponents," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2568–2573, Sep. 2002.

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

[10] I. Herstein, *Topics In Algebra*, 2nd ed. New York, NY, USA: Wiley, 1975.

[11] N. Merhav, "Error exponents of typical random codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6223–6235, Sep. 2018.

[12] D. de Caen, "A lower bound on the probability of a union," *Discrete Math.*, vol. 169, nos. 1–3, pp. 217–220, May 1997.

[13] R. J. McEliece, E. R. Rodemich, H. Rumsey, and L. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 2, pp. 157–166, Mar. 1977.

[14] S. Litsyn, "New upper bounds on error exponents," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 385–398, Mar. 1999.

[15] G. D. Forney, Jr., "Generalized minimum distance decoding," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 125–131, Apr. 1966.

Anshoo Tandon (S'13–M'17) received the B.E. degree in computer science and engineering from Kumaun University, Nainital, India, in 1998, the M.E. degree in signal processing from the Indian Institute of Science, Bengaluru, India, in 2000, and the Ph.D. degree from the National University of Singapore (NUS), Singapore, in 2016.

From 2000 to 2011, he worked in different capacities in the industry toward developing efficient cellular and wireless connectivity solutions. He is currently a Research Fellow with the Department of Electrical and Computer Engineering, NUS. His research interests include information and coding theory, and algebra.

Vincent Y. F. Tan (S'07–M'11–SM'15) was born in Singapore, in 1981. He received the B.A. and M.Eng. degrees in electrical and information sciences from Cambridge University in 2005, and the Ph.D. degree in electrical engineering and computer science (EECS) from the Massachusetts Institute of Technology (MIT) in 2011.

He is currently a Dean's Chair Associate Professor with the Department of Electrical and Computer Engineering and the Department of Mathematics, National University of Singapore (NUS). He has authored a research monograph titled *Asymptotic Estimates in Information Theory With Non-Vanishing Error Probabilities* in the Foundations and Trends in Communications and Information Theory Series (NOW Publishers). His research interests include information theory, machine learning, and statistical signal processing.

Dr. Tan received the MIT EECS Jin-Au Kong Outstanding Doctoral Thesis Prize in 2011, the NUS Young Investigator Award in 2014, the Singapore National Research Foundation (NRF) Fellowship (Class of 2018), and the NUS Young Researcher Award in 2019. He is also an IEEE Information Theory Society Distinguished Lecturer in 2018 and 2019. He is also serving as an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Lav R. Varshney (S'00–M'10–SM'15) received the B.S. degree *(magna cum laude)* with honors in electrical and computer engineering from Cornell University, Ithaca, NY, in 2004, and the S.M., E.E., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 2006, 2008, and 2010, respectively.

From 2010 to 2013, he was a Research Staff Member with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY. He is currently an Assistant Professor with the Coordinated Science Laboratory and also with the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign. Since 2019, he has been on leave as a Principal Research Scientist with Salesforce Research, Palo Alto, CA. His research interests include information and coding theory, statistical signal processing, neuroscience, and artificial intelligence.

Dr. Varshney is also a member of Eta Kappa Nu, Tau Beta Pi, and Sigma Xi. He was a recipient of IBM Faculty Award in 2014 and was a Finalist for the Bell Labs Prize in 2014 and 2016. He and his students have received several best paper awards. He also serves on the advisory board for the AI XPRIZE.