The Importance of Tie-Breaking in Finite-Blocklength Bounds

Eli Haim
Dept. of EE-Systems, TAU
Tel Aviv, Israel
Email: elih@eng.tau.ac.il

Yuval Kochman
School of CSE, HUJI
Jerusalem, Israel
Email: yuvalko@cs.huji.ac.il

Uri Erez
Dept. of EE-Systems, TAU
Tel Aviv, Israel
Email: uri@eng.tau.ac.il

***Abstract*—Upper bounds on the error probability in channel coding are considered, improving the RCU bound by taking into account events, where the likelihood of the correct codeword is tied with that of some competitors. This bound is compared to various previous results, both qualitatively and quantitatively; it is shown to be the tightest bound with respect to previous bounds with the same computational complexity. With respect to maximal error probability of linear codes, it is observed that when the channel is additive, the derivation of bounds, as well as the assumptions on the admissible encoder and decoder, simplify considerably.**

I. Introduction

Consider maximum-likelihood decoding, known to be optimal in the sense of *average* error probability between equiprobable messages. We address the case where a tie occurs, i.e., when ℓ false codewords share the maximum likelihood score with the transmitted one. Regardless of how such a tie is broken, the *average* error probability given this event is equal to $1 - 1/(\ell+1)$. Due to ties, the *complexity* of computing the optimal error probability grows exponentially with the blocklength (see e.g., [1, Theorem 15]), and is cumbersome even for moderate blocklength. For this reason simpler bounds were derived by neglecting ties (see e.g., [1]).

In this paper we study the effect of ignoring this event, i.e., assuming that in case of a tie, the decoder is always right or always wrong. The effect depends upon both the channel and the blocklength. When the likelihood score is a continuous random variable, the probability of ties is zero. Also, for long enough blocks, the distribution of the score of a codeword can be closely approximated by a continuous score (e.g., using the central-limit theorem).

Nonetheless, the effect of ties is important in two respects. First, it may be non-negligible for small enough alphabet size and moderate blocklength. More importantly, as we show in this paper, the way ties are handled is responsible for the counterintuitive phenomenon, that the threshold-based decoding bound (DT) may in certain cases be tighter than the RCU bound, which is based on maximum likelihood (ML) decoding. Specifically, this phenomenon is shown to be an artifact of the way ties are handled in the bounds. We show this by revisiting the finite-blocklength achievability results of Polyanskiy et al. [1]. By carefully considering tie events, we derive a slightly refined RCU bound, that is always tighter than bounds based upon threshold decoding.

In addition, when it comes to maximal error probability, tie-breaking is no longer merely an analysis issue. Rather, ties have to be broken in a balanced manner, such that the error probability for different messages is equal. In [1], a randomized decoder is employed to achieve such fairness. We show that for additive channels (over a finite field), a deterministic decoder suffices.

II. Notation and Background

We consider coding over a memoryless channel with some finite blocklength n, i.e.:

$$V(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} V(y_i|x_i). \tag{1}$$

for $\mathbf{x} \in \mathcal{X}^n, \mathbf{y} \in \mathcal{Y}^n$. The channel input and output alphabets are arbitrary. For the sake of simplicity, we adopt a discrete notation.[1] The codebook is given by $\mathbf{x}_1, \ldots, \mathbf{x}_M$, where M is the number of codewords (that is, the code rate is $R = 1/n \log M$). The decoder produces an estimate $\hat{m}$, where the transmitted message index is denoted by m. The average error probability, assuming equiprobable messages, is given by:

$$\epsilon = \frac{1}{M} \sum_{m=1}^{M} \mathbb{P}\left(\hat{m} \neq m | \mathbf{X} = \mathbf{x}_m\right). \tag{2}$$

The maximum error probability is given by

$$\overline{\epsilon} = \max_{m=1\ldots M} \mathbb{P}\left(\hat{m} \neq m | \mathbf{X} = \mathbf{x}_m\right). \tag{3}$$

For the sake of analyzing the error probability, it is convenient to consider code ensembles. All the ensembles we consider in this paper fall in the following category.

Definition 1 (Random conditionally-symmetric ensemble): An ensemble is called a random conditionally-symmetric ensemble (RCSE) if its codewords are drawn such that for every different $m, j, k \in \{1, \ldots M\}$ and for every $\mathbf{x}, \bar{\mathbf{x}} \in \mathcal{X}^n$:

$$\mathbb{P}\left(\mathbf{X}_j = \bar{\mathbf{x}} | \mathbf{X}_m = \mathbf{x}\right) = \mathbb{P}\left(\mathbf{X}_k = \bar{\mathbf{x}} | \mathbf{X}_m = \mathbf{x}\right). \tag{4}$$

It is easy to verify, that for an RCSE, all codewords are identically distributed. Thus, $\mathbf{X}$ denotes a codeword drawn according

[1] The bounds do not depend on alphabet sizes, and the results can easily be translated the results to the continuous case (which is of limited interest in the context of tie-breaking).

to the (not necessarily memoryless) ensemble distribution P over the set $\mathcal{X}^n$. With this input distribution, the information density of a pair $(\mathbf{x}, \mathbf{y})$ is given by:

$$i(\mathbf{x};\mathbf{y}) = \log \frac{V(\mathbf{y}|\mathbf{x})}{PV(\mathbf{y})}, \tag{5}$$

where $PV(\mathbf{y})$ is the output distribution induced by $P(\mathbf{x})$ and $V(y|x)$. $\mathbf{Y}$ denotes the output corresponding to the random input $\mathbf{X}$, and the random variable $i(\mathbf{X};\mathbf{Y})$ is defined accordingly. In addition, we define a codeword $\bar{\mathbf{X}}$ as a codeword other[2] than the one that generated $\mathbf{Y}$.[3] Therefore, $i(\bar{\mathbf{X}};\mathbf{Y})$ is the information density of a pair $(\bar{\mathbf{X}}, \mathbf{Y})$.

The importance of deriving bounds for an RCSE is due to the fact that this class includes many interesting ensembles. An important special case of RCSE is the pairwise-independent ensemble:

Definition 2 (Pairwise-independent ensemble): A pairwise independent ensemble (PIE) is an ensemble such that its codewords are pairwise-independent and identically distributed. That is, for any two indices $i \neq j$ and an index m,

$$\mathbb{P}\left(\mathbf{X}_i = \mathbf{x}_i | \mathbf{X}_j = \mathbf{x}_j\right) = \mathbb{P}\left(\mathbf{X}_i = \mathbf{x}_i\right) = P(\mathbf{x}_m). \tag{6}$$

We note that the codewords of an RCSE are not necessarily pairwise independent. One example is of linear random codes with a cyclic generating matrix [2]. Generally, an RCSE (which are not necessarily PIE) can be constructed by first drawing a class of codewords, and then, randomly (uniformly and independently) drawing the codewords from this class.

Finally, the following class of channels turns out to play a special role.

Definition 3 (Additive channels): A channel is additive over a finite group $\mathcal{G}$ with an operation, if $\mathcal{X} = \mathcal{Y} = \mathcal{G}$, and the transition distribution $V(y|x)$ is compatible with

$$Y = X + N$$

where N is statistically independent of X, and "$+$" denotes the operation over $\mathcal{G}$.[4]

For example, for discrete modulo-additive channels over $\mathbb{Z}_q$, addition is performed modulo q. The importance of additive channels stems from the following lemma.

Lemma 1: Consider an additive channel over $\mathcal{G}$, and a codebook drawn from a PIE with uniform input distribution over $\mathcal{G}^n$, i.e. $P(\mathbf{x}) = |\mathcal{G}|^{-n} \quad \forall \mathbf{x} \in \mathcal{G}^n$. Then, $i(\bar{\mathbf{X}};\mathbf{Y})$ is statistically independent of $(\mathbf{X}, \mathbf{Y})$.

Proof: For this channel the information density (5) of $(\bar{\mathbf{x}}, \mathbf{y})$ is equal to

$$i(\bar{\mathbf{x}};\mathbf{y}) = \log \frac{P_{\mathbf{N}}(\mathbf{y}-\bar{\mathbf{x}})}{P_{\mathbf{Y}}(\mathbf{y})}, \tag{7}$$

where $P_{\mathbf{N}}(\cdot)$ is the noise distribution, and $P_{\mathbf{Y}}(\cdot)$ is the channel output distribution. Since $\bar{\mathbf{X}}$ is uniformly distributed over $\mathcal{G}^n$ and statistically independent of $(\mathbf{X}, \mathbf{N})$, then $\mathbf{Y} - \bar{\mathbf{X}}$ is statistically independent of $(\mathbf{X}, \mathbf{Y})$; moreover, any function of $\mathbf{Y} - \bar{\mathbf{X}}$ is also statistically independent of $(\mathbf{X}, \mathbf{Y})$. In particular $P_N(\mathbf{Y} - \bar{\mathbf{X}})$ is independent of $(\mathbf{X}, \mathbf{Y})$.

Since $\mathbf{X}$ is uniformly distributed over $\mathcal{G}^n$, the channel output $\mathbf{Y}$ is also uniformly distributed over $\mathcal{G}^n$. Hence, $P_{\mathbf{Y}}(\mathbf{Y})$ is statistically independent of $(\mathbf{X}, \mathbf{Y})$. These two observations conclude the proof. ■

III. I.I.D. Codebooks

Before stating the main results that apply to any RCSE, we begin with simple bounds that hold for the special case of an i.i.d. ensemble. That is, all codewords are mutually independent, and each is distributed according to $P(\mathbf{X})$. In this case, the average error probability is well known, although hard to compute [1]. Denote:

$$W = \mathbb{P}\left(i(\bar{\mathbf{X}};\mathbf{Y}) = i(\mathbf{X};\mathbf{Y}) | \mathbf{X}, \mathbf{Y}\right) \tag{8a}$$

$$Z = \mathbb{P}\left(i(\bar{\mathbf{X}};\mathbf{Y}) < i(\mathbf{X};\mathbf{Y}) | \mathbf{X}, \mathbf{Y}\right). \tag{8b}$$

Then, for an i.i.d. ensemble [1, Thm. 15]:

$$\epsilon^{(\text{iid})} = 1 - \sum_{\ell=0}^{M-1} \frac{1}{\ell+1} \cdot \binom{M-1}{\ell} \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left(W^{\ell} Z^{M-1-\ell}\right). \tag{9}$$

This result follows since for equiprobable codewords, ML decoding amounts to maximum information density. We note that ℓ represents the number of competing codewords that share the maximal information density score with the correct one; given ℓ, the correct codeword will be chosen with probability $1/(\ell+1)$. If $W = 0$ (as happens when $V(Y|x)$ is a proper density for every $x \in \mathcal{X}$), the calculation is straightforward. Otherwise, it has exponential complexity. Thus, the main burden is with dealing with ties. The following simple bounds can be used to circumvent this burden.

Proposition 1 (Bounds for i.i.d. codebooks): For an i.i.d. ensemble,

$$1 - \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[(W+Z)^{M-1}\right] \leq \epsilon^{(\text{iid})} \leq 1 - \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[Z^{M-1}\right]. \tag{10}$$

This result can be shown either from (9) or directly. For the lower bound, in case multiple codewords (including the correct one) attain the maximal information density, the correct one is always chosen; for the upper bound, it is never chosen under such circumstances. Of course, as the upper bound is just the first term in (9), one may tighten it by taking more terms. The difference between the lower and upper bounds may be quite significant, as demonstrated in Figure 1.

IV. Bounds for RCSE

A. Maximum-Likelihood Union Bounds

When the codewords are not statistically independent, we can no longer use products of probabilities as done in the previous section. However, for providing an upper bound on the error probability, we can use a union bound. We derive a

[2] In a random codebook it may happen that the codebook contains some identical codewords. Thus it is possible that $\bar{\mathbf{X}} = \mathbf{X}$, as long as they represent different messages.

[3] In [1], the notation $i(\mathbf{X};\bar{\mathbf{Y}})$ is sometimes used; for RCSE, the two are equivalent.

[4] The operation "$-$" over the group, which is uniquely defined by the operation "$+$", such that for any $a, b, c \in \mathcal{G}$: $a - b = c$ iff $a = c + b$.

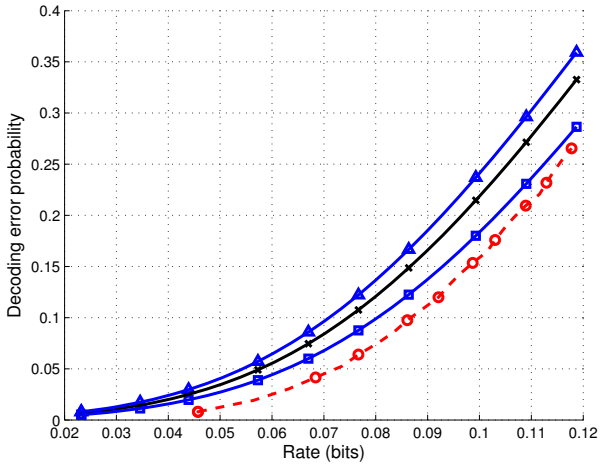

Fig. 1. The effect of tie-breaking on the performance of i.i.d. codebooks. We demonstrate the effect using a BSC with crossover probability 0.3, at blocklength $n = 100$. The triangle- and square- marked solid curves give the lower and upper bounds of Proposition 1, respectively. The $\times$-marked solid curve is the exact error probability of the i.i.d. ensemble (9), evaluated by taking enough terms in the sum, such that the effect of additional ones is numerically insignificant. For reference, the circle-marked dashed curve gives the tightest lower bound on the error probability, which holds for *any* codebook [1, Theorem 16].

result that is close in spirit to the RCU bound [1, Theorem 16], which states that $\epsilon^{(\text{iid})} \leq \epsilon_{\text{RCU}}$,[5] where

$$\epsilon_{\text{RCU}} \triangleq \mathbb{E}_{\mathbf{X},\mathbf{Y}} \left[\min \left\{ 1, (M-1) \cdot (1 - Z) \right\} \right]. \tag{11}$$

We improve this bound in two ways: first, it is extended to any RCSE, and second, the case of equal maximal information-density scores (ties) is taken into account.

Theorem 1 (RCU* bound): The average error probability of an RCSE satisfies $\epsilon^{(\text{RCSE})} \leq \epsilon_{\text{RCU}^*}$ where

$$\epsilon_{\text{RCU}^*} \triangleq \mathbb{E}_{\mathbf{X},\mathbf{Y}} \left[\min \left\{ 1, (M-1) \cdot \left(1 - Z - \frac{W}{2} \right) \right\} \right], \tag{12}$$

where the conditional probabilities W and Z are given by (8).

Proof: Without loss of generality, assume that the transmitted codeword index is $m = 1$. The ML decoder will choose the codeword with maximal information density; in case of equality, it will uniformly draw a "winner" between the maximal ones. Let C_j be the event that the codeword j was chosen in such a drawing. Denote the following events:

$$A_j \triangleq \{ i(\mathbf{X}_j; \mathbf{Y}) > i(\mathbf{X}; \mathbf{Y}) \} \tag{13a}$$

$$B_j \triangleq \{ i(\mathbf{X}_j; \mathbf{Y}) = i(\mathbf{X}; \mathbf{Y}) \}. \tag{13b}$$

and $A \triangleq \bigcup_{j=2}^{M} A_j$, $B \triangleq \bigcup_{j=2}^{M} B_j$. Also denote

$$\Psi \triangleq \{ m = 1 \wedge \mathbf{X}_1 = \mathbf{x}_1 \wedge \mathbf{Y} = \mathbf{y} \}. \tag{14}$$

[5]Indeed, it is noted in [1, Appendix A] that pairwise independence is sufficient.

Then

$$\begin{aligned} S(\mathbf{x}_1, \mathbf{y}) &\triangleq \mathbb{P}\left(A \cup [B \cap {}^{c}C_1] | \Psi \right) && \text{(15a)} \\ &\leq \mathbb{P}\left(A | \Psi \right) + \mathbb{P}\left(B \cap {}^{c}C_1 | \Psi \right) && \text{(15b)} \\ &= \mathbb{P}\left(A | \Psi \right) + \mathbb{P}\left(B | \Psi \right) \cdot \mathbb{P}\left({}^{c}C_1 | \Psi, B \right) && \text{(15c)} \\ &\leq \mathbb{P}\left(A | \Psi \right) + \frac{1}{2} \mathbb{P}\left(B | \Psi \right) && \text{(15d)} \\ &\leq (M-1)(1 - W - Z) + \frac{1}{2}(M-1)W && \text{(15e)} \\ &= (M-1)\left(1 - Z - W/2\right), && \text{(15f)} \end{aligned}$$

where (15b) is due to the union bound; (15d) holds since given that there is at least one competing codeword (any codeword other than the transmitted one, having maximal score), then the probability that the transmitted codeword will not be chosen in the decoding is less than $1/2$; (15e) is due to the union bound.

The error probability is given by:

$$\begin{aligned} \epsilon^{(RCSE)} &= \mathbb{P}\left(A \cup [B \cap {}^{c}C_1] | m = 1 \right) && \text{(16a)} \\ &= \mathbb{E}_{\mathbf{X}_1,\mathbf{Y}} S(\mathbf{X}_1, \mathbf{Y}) && \text{(16b)} \\ &= \mathbb{E}_{\mathbf{X}_1,\mathbf{Y}} \min\left\{ 1, S(\mathbf{X}_1, \mathbf{Y}) \right\}. && \text{(16c)} \end{aligned}$$

Substituting (15f) in (16c) concludes the proof. ■

Remark 1: We can give the RCU* bound the following interpretation. First, each potential input $\mathbf{x}_j$ is given an information-density score (equivalent to a likelihood score) i_j. Then, these scores are fed to a comparison process. The process is biased against the correct codeword, in the sense that it has to beat each and every competing codeword. However, each pairwise comparison itself is optimal (the correct codeword will beat a competing one with lower score), and fair (in case of a tie, both codewords are equally likely to win). This comparison mechanism is worse than the actual decoder used in the proof, since in the case where the correct codeword shares the maximal score with ℓ competitors, it has probability $2^{-\ell}$ to be chosen, rather than $1/(\ell+1)$; yet, the union bound for both is equal.

B. Relation to Gallager's Type-I bound

The following bound is due to Gallager.

Proposition 2 (Gallager type-I bound [3, Sec. 3.3]): For any constant t

$$\epsilon^{(RCSE)} \leq \epsilon_{\text{G-I}}(t), \tag{17}$$

where

$$\begin{aligned} \epsilon_{\text{G-I}}(t) \triangleq{}& \mathbb{P}\left(i(\mathbf{X}; \mathbf{Y}) < t \right) + \\ &+ (M-1) \mathbb{P}\left(i(\mathbf{X}; \mathbf{Y}) \geq t \wedge i(\bar{\mathbf{X}}; \mathbf{Y}) \geq i(\mathbf{X}; \mathbf{Y}) \right). \end{aligned} \tag{18}$$

Just like the RCU, this bound is obtained by applying a union bound for the ML decoder. However, it is inferior to the RCU bound, due to the following consideration. Taking the minimum between the union and 1 in the RCU bound is similar to the threshold t in (18), in the way that it avoids over-estimating the error probability in cases where the channel behavior was "bad". However, whereas the RCU bound uses

the optimal threshold given $\mathbf{X}$ and $\mathbf{Y}$, the Gallager bound uses a *global* threshold, which reflects a tradeoff. Nevertheless, for additive channels (recall Definition 3) the local and global thresholds coincide.

Proposition 3: For any RCSE and for any t:

$$\epsilon_{\text{G-I}}(t) \geq \epsilon_{\text{RCU}}, \tag{19}$$

where $\epsilon_{\text{G-I}}(t)$ and ϵ_{RCU} are defined in (18) and in (11) respectively. If the channel is additive and the code ensemble is PIE with uniform distribution over $\mathcal{X}$, then there exists a value t such that equality holds.

Proof: For the first part, define the events $A \triangleq \left\{i(\bar{\mathbf{X}};\mathbf{Y}) \geq i(\mathbf{X};\mathbf{Y})\right\}$ and $T \triangleq \{i(\mathbf{X};\mathbf{Y}) \geq t\}$ (^cT denotes the complementary event). Then:

$$\begin{aligned}
\epsilon_{\text{RCU}} &= \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,(M-1)\cdot(1-Z)\right\}\right] && \text{(20a)}\\
&\leq \mathbb{P}\left({}^cT\right) + \mathbb{P}\left(T\right)\cdot\mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[(M-1)\mathbb{P}\left(A|\mathbf{X},\mathbf{Y}\right)|T\right] && \text{(20b)}\\
&= \mathbb{P}\left({}^cT\right) + (M-1)\mathbb{P}\left(T\right)\cdot\mathbb{P}\left(A|T\right) && \text{(20c)}\\
&= \epsilon_{\text{G-I}}(t) && \text{(20d)}
\end{aligned}$$

For the second part, recall that by Lemma 1, $i(\bar{\mathbf{X}};\mathbf{Y})$ is statistically independent of $(\mathbf{X},\mathbf{Y})$. Denote by t^* the minimal threshold t such that

$$(M-1)\mathbb{P}\left(i(\bar{\mathbf{X}};\mathbf{Y}) \geq t\right) \leq 1.$$

Then $(M-1)\mathbb{P}\left(i(\bar{\mathbf{X}};\mathbf{Y}) \geq i(\mathbf{X};\mathbf{Y})\middle|i(\mathbf{X};\mathbf{Y}) < t^*\right) \geq 1$. We have that: $\mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,(M-1)(1-Z)\right\}|i(\mathbf{X};\mathbf{Y}) < t^*\right] = 1$, i.e., the inequality in (20b) becomes an equality in this case. ∎

Remark 2: It follows, that for the BSC, $\epsilon_{\text{G-I}}(t^*) = \epsilon_{\text{RCU}}$. Indeed, it is noted in [1] that for the BSC, the RCU bound is equal to Poltyrev's bound [4]; this is not surprising, since the latter is derived from (18) (Poltyrev's bound uses linear codes, see Section V in the sequel).

Remark 3: Gallager's type I bound can be improved by breaking ties, similar to the improvement of RCU*, leading to G-I*. An analogous result to Proposition 3 relates G-I* and RCU*.

C. Threshold-Decoding Union Bounds

The average error probability of an RCSE can be further upper-bounded using the sub-optimal *threshold decoder* [5]. This decoder looks for a codeword that has a likelihood score above some predetermined threshold. In [1, Theorem 18] a union bound is derived for such a decoder, where if multiple codewords pass the threshold, the winner is chosen uniformly from among them.[6] The resulting "dependence testing" (DT) bound is given by:

$$\epsilon_{\text{DT}}(t) \triangleq \mathbb{P}\left(i(\mathbf{X};\mathbf{Y}) \leq t\right) + \frac{M-1}{2}\mathbb{P}\left(i(\bar{\mathbf{X}},\mathbf{Y}) > t\right), \tag{21a}$$

where the optimal threshold is given by[7]

$$t = \log\frac{M-1}{2}. \tag{21b}$$

A troubling behavior, demonstrated in [1] using the binary erasure channel (BEC), is that sometimes $\epsilon_{\text{RCU}} > \epsilon_{\text{DT}}$. This is counterintuitive since the DT bound is derived by applying a union bound to a sub-optimal decoder. We find that this phenomenon stems from the fact that the RCU bound ignores ties, and prove that the improved bound, denoted by RCU*, always satisfies $\epsilon_{\text{RCU}^*} \leq \epsilon_{\text{DT}}$. To that end, we also prove a (very slightly) improved bound for the threshold decoder, which will be referred as the TU bound, that is closer in form to the ML bounds (11) and (18) and allows for a simpler comparison. It uses the following definitions (cf. (8)).

Let

$$W_q = \mathbb{P}\left(q(i(\bar{\mathbf{X}};\mathbf{Y})) = q(i(\mathbf{X};\mathbf{Y}))|\mathbf{X},\mathbf{Y}\right) \tag{22a}$$

$$Z_q = \mathbb{P}\left(q(i(\bar{\mathbf{X}};\mathbf{Y})) < q(i(\mathbf{X};\mathbf{Y}))|\mathbf{X},\mathbf{Y}\right), \tag{22b}$$

where $q(i)$ is the indicator function:

$$q(i) \triangleq \mathbb{1}_{\{i>t\}}. \tag{22c}$$

Proposition 4: For an RCSE and for any t,

$$\epsilon^{(\text{RCSE})} \leq \epsilon_{\text{TU}}(t), \tag{23}$$

where

$$\epsilon_{\text{TU}}(t) \triangleq \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,(M-1)\cdot\left(1-Z_q-\frac{W_q}{2}\right)\right\}\right]. \tag{24}$$

Furthermore, $\epsilon_{\text{TU}}(t) \leq \epsilon_{\text{DT}}(t)$.

Proof: For proving achievability, consider a decoder identical to the ML decoder, except that before comparing the codewords, the information-density scores are quantized according to (22c). For the comparison to the DT bound, denote $Q \triangleq \mathbb{P}\left(i(\bar{\mathbf{X}},\mathbf{Y}) > t\middle|\mathbf{X},\mathbf{Y}\right)$. Then:

$$\begin{aligned}
\epsilon_{\text{DT}}(t) &= E_{\mathbf{X},\mathbf{Y}}\left[\mathbb{1}_{\{i(\mathbf{X};\mathbf{Y})\leq t\}} + \frac{M-1}{2}Q\right]\\
&\geq E_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,(M-1)\cdot\left[\mathbb{1}_{\{i(\mathbf{X};\mathbf{Y})\leq t\}} + \frac{1}{2}Q\right]\right\}\right]\\
&\geq E_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,\frac{M-1}{2}\cdot\left[\mathbb{1}_{\{i(\mathbf{X};\mathbf{Y})\leq t\}} + Q\right]\right\}\right]\\
&= \epsilon_{\text{TU}}(t).
\end{aligned}$$

∎

Remark 4: It is not obvious that the optimal threshold for the TU bound is t of (21b). However, it is good enough for our purposes.

Proposition 5: For any channel, $\epsilon_{\text{RCU}^*} \leq \epsilon_{\text{TU}}(t)$ for any t. Thus, the RCU* bound is tighter than the DT bound, i.e.:

$$\epsilon_{\text{RCU}^*} \leq \epsilon_{\text{DT}}(t).$$

[6] In fact, the proof states that the "first" codeword to pass the threshold is selected. However, such ordering of the codewords is not required.

[7] In [6], the threshold is further optimized, depending on the competing codeword and on the received word

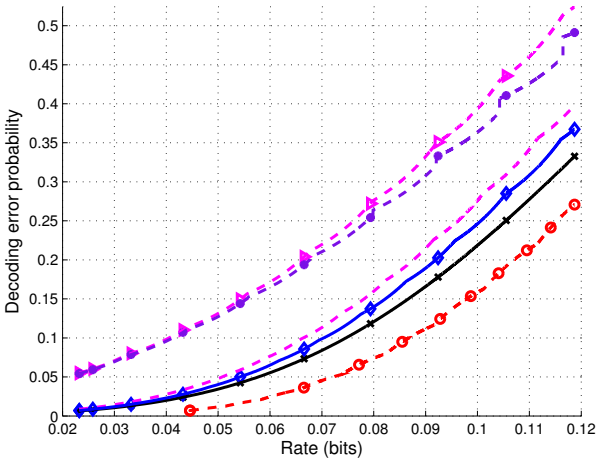

Fig. 2. The effect of tie-breaking on the performance of PIE codebooks of the different union bounds. We demonstrate the effect using a BSC with crossover probability 0.3, at blocklength $n = 100$. The triangle-marked dashed curve is the DT bound (21). The asterisk-marked dashed curve is the TU bound (24). The dashed curve is the RCU bound (11). The diamond-marked solid curve is the RCU* bound (12). For reference, we repeat two of the curves of Figure 1. The $\times$-marked solid curve is the exact performance of the i.i.d. ensemble (9), while the circle-marked dashed curve is the lower bound for any codebook [1, Theorem 16]. The non-smoothness of some of the curves is not an artifact, but comes from the fact that they involve integers.

Proof: Recalling Remark 1, the RCU* bound reflects optimal (ML) pairwise decision. Thus, necessarily the pairwise error probabilities satisfy $Z + W/2 \geq Z_q + W_q/2$. ■

Remark 5: The case of the BEC, where $\epsilon_{\text{RCU}^*} = \epsilon_{\text{TU}} = \min(1, \epsilon_{\text{DT}})$ is very special. In the BEC, a competing codeword cannot have a higher score than the true codeword; if the channel realization is such that the non-erased elements of $\mathbf{x}$ and $\bar{\mathbf{x}}$ are equal, then $i(\bar{\mathbf{x}}; \mathbf{y}) = i(\mathbf{x}; \mathbf{y})$, otherwise $i(\bar{\mathbf{x}}; \mathbf{y}) = -\infty$. Thus, $\epsilon_{\text{RCU}^*} = \mathbb{E}_{\mathbf{X},\mathbf{Y}}[\min\{1, (M-1)W/2\}]$. Let k be the number of non-erased symbols out of the block of n. Then $W = 2^{-k}$. Consequently, $(M-1)W/2 > 1$ if and only if $i(\mathbf{x}; \mathbf{y}) < t$, where t is given by (21b) (with logarithms to base 2).

D. Performance Comparison

Comparison of the different union bounds is given in Figure 2. In particular, the effect of tie-breaking on the bounds is shown by the comparison of the RCU bound (11) and the RCU* bound (12). Notice that this bound depends on the ensemble. Due to Lemma 1, the computation of the RCU and RCU* bounds for PIE becomes simple. For this reason bounds are shown for this ensemble. Since an i.i.d. ensemble is also PIE, the exact error probability for i.i.d. ensemble (9) is given as a reference.

V. Linear Codes

A. The Dithered Linear Codes Ensemble

For any finite field $\mathbb{F}_q$ of cardinality q, we define the dithered linear codes ensemble by

$$\mathcal{C} = \left\{ \mathbf{X}_j = G\mathbf{w}_j + \mathbf{D} \middle| \mathbf{w}_j \in \mathbb{F}_q^k \right\}, \tag{25}$$

where G is a $n \times k$ random generator matrix, $\mathbf{D}$ is an n-dimensional random shift, and the elements of G and $\mathbf{D}$ are drawn uniformly over $\mathbb{F}_q$ and are statistically independent. It follows that the codebook size is $M = q^k$.

B. Additive Channels

It is proven in [1, Appendix A] that for a class of channels, which includes the BSC and the BEC, there exists a *randomized* ML decoder such that the *maximal* error probability $\overline{\epsilon}$ (3) coincides with the average one.

We now restrict our attention to channels that are additive, in the sense of Definition 3. Further, assume that the channels are additive over a finite field, which is the same field over which the code is linear. Clearly, in this situation the dither does not change the distance profile of the codebook. Thus, it suffices to consider the linear codes ensemble

$$\mathcal{C} = \left\{ \mathbf{X}_j = G\mathbf{w}_j \middle| \mathbf{w}_j \in \mathbb{F}_q^k \right\}, \tag{26}$$

where again G is i.i.d. uniform over $\mathbb{F}_q$. More importantly, we show that in order to achieve good maximal error probability, there is no need to use randomized decoders.

Theorem 2: For any channel that is additive over a finite field, for an ensemble of linear codes over the field, there exists a deterministic decoder satisfying:

$$\overline{\epsilon} \leq \epsilon_{\text{RCU}^*}. \tag{27}$$

Proof outline: Let $\Omega_1, \ldots, \Omega_M$ be a partition of the output space $\mathcal{Y}^n$ into decision regions (for any $1 \leq m \leq M$, Ω_m is associated with codeword m). From (7), a partition is optimal in the average error probability sense, if and only if it satisfies:

$$\Omega_m \subseteq \left\{ \mathbf{y} \in \mathbb{F}_q^n \middle| \forall m' \neq m : P_{\mathbf{N}}(\mathbf{y} - \mathbf{x}_m) \geq P_{\mathbf{N}}(\mathbf{y} - \mathbf{x}_{m'}) \right\} \tag{28a}$$

$$\Omega_m \supseteq \left\{ \mathbf{y} \in \mathbb{F}_q^n \middle| \forall m' \neq m : P_{\mathbf{N}}(\mathbf{y} - \mathbf{x}_m) > P_{\mathbf{N}}(\mathbf{y} - \mathbf{x}_{m'}) \right\}, \tag{28b}$$

and for all $m \neq m'$: $\Omega_m \cap \Omega_{m'} = \emptyset$. Since for any such optimal partition $\epsilon \leq \epsilon_{\text{RCU}^*}$, it is sufficient to show that there exists a partition satisfying (28) for which $\overline{\epsilon} = \epsilon$. By performing coset decomposition $\Omega_1, \ldots, \Omega_M$ of $\mathbb{F}_q^n$ such that (28) holds, the decision regions are equal up to a translation. Thus, the ties are broken in a balanced manner. ■

References

[1] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[2] G. Seguin, "Linear ensembles of codes," *IEEE Trans. Information Theory*, vol. 25, no. 4, pp. 477 – 480, July 1979.

[3] R. G. Gallager, "Low-density parity-check codes," Ph.D. dissertation, Massachusetts Institute of Technology, 1963.

[4] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. Information Theory*, vol. 40, no. 4, pp. 1284 –1292, July 1994.

[5] A. Feinstein, "A new basic theorem of information theory," *Information Theory, Transactions of the IRE Professional Group on*, vol. 4, no. 4, pp. 2 –22, September 1954.

[6] A. Martinez and A. Guillén i Fàbregas, "Random-coding bounds for threshold decoders: Error exponent and saddlepoint approximation," in *Proc. Int. Symp. Info. Theory*, Aug. 2011.