

Unequal Error Protection Coding Approaches to the Noisy 20 Questions Problem

Hye Won Chung

Lizhong Zheng

Brian M. Sadler

Alfred O. Hero

Abstract—In this paper, we propose an unequal error protection coding strategy based on superposition coding for the noisy 20 questions problem. In this problem, a player wishes to successively refine an estimate of the value of a continuous random variable by posing binary queries and receiving noisy responses. When the queries are designed non-adaptively as a single block and the noisy responses are modeled as the output of a binary symmetric channel the 20 questions problem can be mapped to an equivalent problem of channel coding with unequal error protection (UEP). A superposition coding strategy with UEP is introduced that has error exponent that is significantly better than that of the UEP repetition code introduced by Variani *et al.* [1].

I. INTRODUCTION AND PROBLEM STATEMENT

Classical information theory assumes that each bit of information is equally important. When information bits are transmitted through a noisy channel to a destination, the communication system provides uniform error protection to all transmitted information bits. On the other hand, in many applications, information can be very heterogeneous in its value to the user at the destination. This occurs, for example, when the transmitted information bits correspond to coefficients in the dyadic expansion of a signal that is to be recovered with small mean squared error.

This motivates a value-centered information transmission/extraction strategy where the bits of higher significance get higher levels of error protection.

In this paper, we consider value-centered information extraction in the context of the noisy 20 questions problem [2]. The objective of this problem is to determine the state of a source by querying about the first k bits in the dyadic expansion of $X \sim \text{unif}[0, 1]$. This general setup of noisy 20 questions game has many applications. For example, a target localization problem can be modeled as a noisy 20 questions game where a player aims to locate a target by receiving query responses from sensors probing the region of interest. The sequence of questions is controlled by a controller that may either operate open-loop (non-adaptive 20 questions) or use feedback (adaptive 20 questions).

When queries can be updated based on the collected answers by feeding back these answers to the controller, the querying strategy is called adaptive. For adaptive sequential querying,

Hye Won Chung (hyechung@umich.edu) and Alfred O. Hero (hero@eecs.umich.edu) are with the EECS department at the University of Michigan; Lizhong Zheng (lizhong@mit.edu) is with the EECS department at MIT; Brian M. Sadler (brian.m.sadler6.civ@mail.mil) is with the US Army Research Laboratory. This research was partially supported by the US Army Research Laboratory grant W911NF-11-1-0391.

successive entropy minimization is often proposed as a way to design a sequence of queries. As an example, the bisection policy [3] extracts the maximum 1 bit of information about X at each round by asking whether X lies to the left or right of the median of the posterior distribution, which is updated based on the collected answers. As demonstrated in [4], this policy achieves mean squared error (MSE) that is decreasing exponentially in the number of queries N .

On the other hand, in non-adaptive querying, the sequence of N queries is determined in advance, without the benefit of feedback. For this case, we aim to estimate the first k bits in the dyadic expansion of $X \sim \text{unif}[0, 1]$, forming an estimator of the k -bit quantized X . When the objective is to minimize the MSE of this estimator of X , importance of the quantized bits of X is higher for most significant bits (MSB) than it is for least significant bits (LSB). Thus, any non-adaptive controller should use a querying strategy that can provide unequal error protection (UEP) for each of the information bits.

One way to provide unequal error protection is to repeat the query on the quantized information bits multiple times, the number of repetitions varying in accordance with the desired level of unequal protection. Such a UEP repetition coding approach was considered in [1]. It was shown that the resulting MSE decreases exponentially in \sqrt{N} , which is smaller than the N exponential rate of decrease achieved in the bisection-based adaptive 20 questions [3].

The main contribution of this paper is to provide a new non-adaptive querying strategy based on superposition coding that can provide UEP and achieves better convergence rates for the estimation of X . With the proposed querying strategy, quantized MSE decreases exponentially in N , as contrasted to \sqrt{N} , matching the rate of the adaptive scheme. Furthermore, the scale factor in the exponent is better than that of a random block code employing equal error protection.

A. Problem Statement

To estimate an unknown random variable $X \sim \text{unif}[0, 1]$, a player asks to an oracle whether X is located within some sub-region $Q_i \subset [0, 1]$, either connected or non-connected. The oracle gives the correct answer $Z_i(X) = \mathbb{1}(X \in Q_i)$, and the player receives a noisy version $Y_i \in \{0, 1\}$ of the oracle's answer transmitted through a BSC(ϵ). After receiving (Y_1, \dots, Y_N) , the player calculates an estimate \hat{X}_N of X .

For non-adaptive block querying, the player estimates X by querying about the first k bits in the dyadic expansion of $X \approx 0.B_1 \dots B_k$ where $k = NR/\ln 2$ for a fixed rate

$R > 0$. Discovering (B_1, \dots, B_k) is equivalent to finding the index $M = \sum_{i=1}^k B_i 2^{k-i} \in \{0, \dots, 2^k - 1\}$ of the interval $I_M := [M2^{-k}, (M+1)2^{-k})$ where the value of the target variable belongs, $X \in I_M$. Here the region of interest $[0, 1]$ is uniformly quantized into 2^k disjoint sub-intervals $\{I_0, \dots, I_{2^k-1}\}$ of length 2^{-k} .

The oracle and the player agree on a block of questions (Q_1, \dots, Q_N) , and the oracle provides a block of binary answers (Z_1, \dots, Z_N) . Assume that each querying region Q_i is a union of some subset of quantized intervals $\{I_0, \dots, I_{2^k-1}\}$. By considering the index $M \in \{0, \dots, 2^k - 1\}$ as a message transmitted from the oracle to the player and the oracle's answers as a codeword, block querying can be mapped to an equivalent problem of block channel coding over a BSC(ϵ). Designing a block of questions to discover the index M can be thought of as designing a length- N and rate- R block code. When $\mathbf{z}^{(m)} = (z_1^{(m)}, \dots, z_N^{(m)})$ is a length- N codeword for a message $m \in \{0, \dots, 2^k - 1\}$, the associated i -th querying region Q_i becomes the union of the sub-intervals $\{I_{m'}\}$ for message m' 's such that the i -th answer bit $z_i^{(m')}$ equals 1. Therefore, the encoder specifies a block of questions, and vice versa. After receiving N channel outputs, the player generates estimates $(\hat{B}_1, \dots, \hat{B}_k)$ of (B_1, \dots, B_k) and $\hat{M} = \sum_{i=1}^k 2^{k-i} \hat{B}_i$ of the message M . Define the finite resolution estimator $\hat{X}_N = \hat{M}e^{-NR} + e^{-NR}/2$.

We mainly consider two types of estimation errors. The first is the mean squared error (MSE) $\mathbb{E}[|X - \hat{X}_N|^2]$. The second is the quantized MSE $\mathbb{E}[c_q(X, \hat{X}_N)]$ where the quantized cost function $c_q(X, \hat{X}_N)$ with 2^k levels is a stepwise function defined as

$$c_q(X, \hat{X}_N) = (se^{-NR})^2, \text{ when} \\ se^{-NR} - \frac{e^{-NR}}{2} < |X - \hat{X}_N| \leq se^{-NR} + \frac{e^{-NR}}{2}, \\ \text{for } s \in \{0, \dots, e^{NR} - 1\}.$$

We consider this cost function when the objective is to estimate the value of X only up to the first k bits in the dyadic expansion of $X \approx 0.B_1 \dots B_k$. Note that the quantized MSE equals 0 if the player correctly decodes the message M . The error increases proportional to the square of $|M - \hat{M}|$.

For a given cost function $c(x, \hat{x}_N)$, the expected estimation error can be written in terms of block decoding events $\{\hat{M} = M\}$ and $\{\hat{M} \neq M\}$ as

$$\mathbb{E}[c(X, \hat{X}_N)] = \Pr(\hat{M} \neq M) \mathbb{E}[c(X, \hat{X}_N) | \hat{M} \neq M] \\ + \Pr(\hat{M} = M) \mathbb{E}[c(X, \hat{X}_N) | \hat{M} = M].$$

With the finite resolution estimator $\hat{X}_N = \hat{M}e^{-NR} + e^{-NR}/2$, the MSE and the quantized MSE can be bounded above as

$$\mathbb{E}[|X - \hat{X}_N|^2] \leq \Pr(\hat{M} \neq M) + (e^{-NR}/2)^2, \quad (1)$$

$$\mathbb{E}[c_q(X, \hat{X}_N)] \leq \Pr(\hat{M} \neq M), \quad (2)$$

by using $\mathbb{E}[|X - \hat{X}_N|^2 | \hat{M} = M] \leq (e^{-NR}/2)^2$ and $\mathbb{E}[c_q(X, \hat{X}_N) | \hat{M} = M] = 0$ for $k = NR/\ln 2$, respectively.

Tighter upper bounds on the two estimation errors can be found by expansion in terms of the bit error probabilities. Noting that $\mathbb{E}[|X - \hat{X}_N|^2 | \hat{B}_i \neq B_i, \hat{B}_1^{i-1} = B_1^{i-1}] \leq 2^{-2(i-1)}$, we can find upper bounds on the MSE and on $\mathbb{E}[c_q(X, \hat{X}_N)]$

$$\mathbb{E}[|X - \hat{X}_N|^2] \leq \sum_{i=1}^k \Pr(\hat{B}_i \neq B_i) 2^{-2(i-1)} + 2^{-2k}, \\ \mathbb{E}[c_q(X, \hat{X}_N)] \leq \sum_{i=1}^k \Pr(\hat{B}_i \neq B_i) 2^{-2(i-1)}, \quad (3)$$

where $k = NR/\ln 2$ for the code of rate R (nats/channel use).

Note that the upper bounds in (3) show how differently each bit error probability affects the estimation errors. As the bit position i increases corresponding to less significance, the weights on the bit error probabilities decrease exponentially in i . In order to optimize the upper bound in (3), we need to design a block code that can provide unequal error protection for the information bits depending on the bit positions.¹

II. RANDOM BLOCK CODE VS. UEP REPETITION CODE

In this section, we review two representative codes, random block codes that provide equal error protection for every information bit, and UEP repetition codes that can provide unequal error protection through time-division approaches.

A. Random Block Code

A random block code of length N and rate R encodes $(NR/\ln 2)$ bits of information in length- N binary codewords, while treating each bit with equal importance. For the random block code of rate R , define the best achievable error exponent as $E_r(R) := \liminf_{N \rightarrow \infty} \frac{-\ln \Pr(\hat{M} \neq M)}{N}$. For a BSC(ϵ) with the optimum input distribution Bernoulli(1/2), Forney's analysis [5] provides a closed form solution for $E_r(R)$,

$$E_r(R) = \begin{cases} E_0(1/2, \epsilon) - R, & 0 \leq R < R_{\text{crit}}, \\ d(\gamma_{\text{GV}}(R) || \epsilon), & R_{\text{crit}} \leq R \leq C, \end{cases} \quad (4)$$

where $E_0(a, b) = -\log(1 - 2a(1-a)(\sqrt{b} - \sqrt{1-b})^2)$, $R_{\text{crit}} = d(\gamma_{\text{crit}}(\epsilon) || 1/2)$ with $\gamma_{\text{crit}}(\epsilon) = \frac{\sqrt{\epsilon}}{\sqrt{\epsilon} + \sqrt{1-\epsilon}}$, $C = h(1/2) - h(\epsilon)$, and $\gamma_{\text{GV}}(R)$ satisfies $d(\gamma_{\text{GV}}(R) || 1/2) = R$.

By plugging this error exponent of $\Pr(\hat{M} \neq M) \doteq e^{-NE_r(R)}$ into the upper bounds on the MSE in (1), we obtain

$$\mathbb{E}[|X - \hat{X}_N|^2] \dot{\leq} e^{-NE_r(R)} + e^{-2NR} \doteq e^{-N \min\{E_r(R), 2R\}} \quad (5)$$

The tightest upper bound on $\mathbb{E}[|X - \hat{X}_N|^2]$ is obtained at the rate R maximizing the exponent $\min\{E_r(R), 2R\}$. As shown in Fig 1, the optimum R^* occurs at the crossing point of $E_r(R)$ and $2R$.

¹Notations: Capital letters represent random variables and lower case letters represent specific realizations of those random variables. The entropy of a binary random variable X distributed as Bernoulli(α) is denoted $h(\alpha)$. The Kullback-Leibler divergence between two Bernoulli distributions Bernoulli(α) and Bernoulli(β) is denoted $d(\alpha || \beta)$. The normalized Gilbert-Varshamov distance $0 \leq \gamma_{\text{GV}}(R) \leq 1/2$ is the value $\gamma_{\text{GV}}(R)$ that gives $d(\gamma_{\text{GV}}(R) || 1/2) = R$. Let the bold face \mathbf{z} or \mathbf{z}_1^N denote the length- N binary sequence $(z_1 z_2 \dots z_N)$ where z_t is the t -th bit of \mathbf{z} . The bit-wise XOR operation is symbolized by \oplus . We use \doteq and $\dot{\leq}$ as follows: $a_N \doteq e^{Nd}$ denotes $d = \liminf_{N \rightarrow \infty} \frac{\ln a_N}{N}$, $a_N \dot{\leq} e^{Nd}$ denotes $d \geq \liminf_{N \rightarrow \infty} \frac{\ln a_N}{N}$.

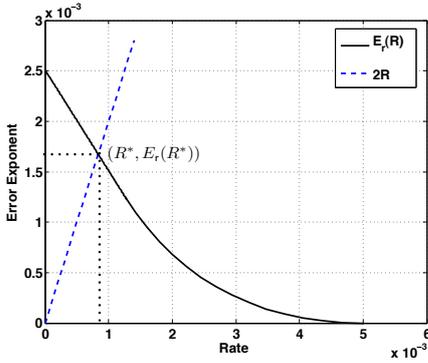


Fig. 1. The random block code error exponent $E_r(R)$ of a BSC(0.45) and the line $2R$; the optimum R^* that maximizes $\min\{E_r(R), 2R\}$ occurs at $R^* \approx C/6$, and the resulting error exponent $E_r(R^*)$ is $0.0017 \approx C/3$.

As shown in [6], for a very noisy channel ($\epsilon \approx 0.5$) the error exponent in (4) can be approximated as

$$E_r(R) \approx \begin{cases} \frac{C}{2} - R, & 0 \leq R < \frac{C}{4}, \\ (\sqrt{C} - \sqrt{R})^2, & \frac{C}{4} \leq R \leq C. \end{cases} \quad (6)$$

By using this, the MSE exponent for the very noisy BSC can be calculated.

Lemma 1: For a BSC(ϵ) with $\epsilon \approx 0.5$, at $R = R^* = \frac{C}{6}$

$$\mathbb{E}[|X - \hat{X}_N|^2] \doteq e^{-N \frac{C}{3}}, \quad (7)$$

where the capacity of the BSC(ϵ), $C = h(1/2) - h(\epsilon)$.

For the quantized estimation error of rate R , we obtain

$$\mathbb{E}[c_q(X, \hat{X}_N)] \leq e^{-N E_r(R)}, \quad (8)$$

whose exponent $E_r(R)$ keeps decreasing as R increases. From the fact that the random block code provides equal protection for every information bit, we can also show that the bound on the MSE in (5) and the bound on the quantized estimation error in (8) are exponentially tight as $N \rightarrow \infty$.

B. UEP Repetition Code

For repetition codes, each raw bit B_i of $M = (B_1, B_2, \dots, B_k)$ is repeatedly transmitted N_i times, where the total number of channel uses is restricted to $\sum_{i=1}^k N_i = N$. In [1], this code was considered in the context of the noisy 20 questions problem. When the information bit B_i is transmitted N_i times through a BSC(ϵ), a simple majority voting algorithm is the maximum likelihood (ML) decoder for B_i achieving

$$\Pr(\hat{B}_i \neq B_i) \leq e^{-N_i d(1/2|\epsilon)}. \quad (9)$$

By assigning different numbers of repetitions (N_1, N_2, \dots, N_k) for each information bit B_i , we can provide unequal error protection. To maximize the exponentially decreasing rate of the bound on the MSE in (3), we need to balance the exponentially decreasing rates of the terms on the right hand side of (3). Since the weight on $\Pr(\hat{B}_i \neq B_i)$ decreases exponentially in i and $\Pr(\hat{B}_i \neq B_i)$ decreases exponentially in N_i , the optimum N_i^* should decrease linearly in i . Therefore, $N = \sum_{i=1}^k N_i^* = O(k^2)$. The optimum k^* is

thus on the order of $O(\sqrt{N})$, and the corresponding optimum rate is $R^* = k^*/N$, which goes to 0 as $N \rightarrow \infty$.

Lemma 2: With the UEP repetition code, we can achieve

$$\mathbb{E}[|X - \hat{X}_N|^2] \leq \left(2\sqrt{\frac{Nd(1/2|\epsilon)}{\ln 2}} + 1 \right) e^{-2\sqrt{(\ln 2)d(1/2|\epsilon)N}}. \quad (10)$$

Compared to the random block code, whose estimation errors decrease exponentially in N , the MSE of the UEP repetition codes decrease exponentially only in \sqrt{N} . Moreover, the UEP repetition code cannot achieve the resolution of e^{-NR} for any $R > 0$, since the code cannot guarantee a reliable communication at any positive rate. To improve the rate of decrease for estimation errors, we need to use more sophisticated codes that have coding gain.

III. UEP WITH SUPERPOSITION CODE

In the previous section, we showed that both unequal error protection and coding gain are necessary in order to improve performance of querying strategies, compared to that of the random block code. In this section, we provide a length- N block querying strategy based on superposition coding that can achieve both UEP and coding gain.

The first k bits in the dyadic expansion of $X \sim \text{unif}[0, 1]$ have different importances in the estimation of X . To minimize the estimation error, it is ideal to provide unequal error protection for each of the information bits. On the other hand, in order to efficiently encode as many bits as possible in a length- N block code, at least some of the information bits should be treated equally. To design a block querying strategy that achieves both UEP and coding gain, we partition the quantized information bits of X into sub-groups, and provide unequal error protection across the groups while treating all the information bits within each group with equal importance.

For the source $X \approx 0.B_1, B_2, \dots, B_k$, the k_1 most significant bits (MSBs) (B_1, \dots, B_{k_1}) of X compose the partial message M_1 and the less significant bits (LSBs) $(B_{k_1+1}, \dots, B_{k_1+k_2})$ of X compose M_2 where $k_1 + k_2 = k$. Denote the rates of M_1 and of M_2 by $R_1 = (k_1 \ln 2)/N$ and $R_2 = (k_2 \ln 2)/N$. Upon transmission of $M = (M_1, M_2)$ of total rate $R_1 + R_2 = R$, the quantized estimation error of resolution e^{-NR} can be bounded as

$$\mathbb{E}[c_q(X, \hat{X}_N)] \leq \Pr(\hat{M}_1 \neq M_1) + \Pr(\hat{M}_2 \neq M_2 | \hat{M}_1 = M_1) e^{-2NR_1}. \quad (11)$$

Note that the weight e^{-2NR_1} on $\Pr(\hat{M}_2 \neq M_2 | \hat{M}_1 = M_1)$ shows that the importance of the partial message M_2 conditioned on the correctly decoded M_1 is significantly smaller than that of M_1 . The question is how to design a coding scheme that provides exponentially better error protection for the more significant partial message M_1 than that of M_2 . With the random block code, the best achievable exponentially decreasing rate of $\Pr(\hat{M}_1 \neq M_1)$ is $E_r(R_1 + R_2)$, and this exponent dominates the convergence rate of $\mathbb{E}[c_q(X, \hat{X}_N)]$. Thus, to improve the estimation error, we need to design a UEP code whose $\Pr(\hat{M}_1 \neq M_1)$ decreases faster than

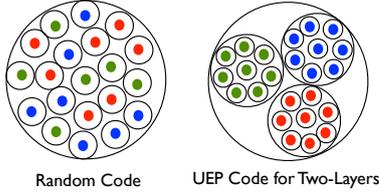


Fig. 2. The distributions of codewords in the output space for random block code and for UEP code with two layers of priority. To better protect color information of the codewords, which represents MSBs of the message of the codewords, the same color codewords should be clustered together.

$e^{-NE_r(R_1+R_2)}$. We show that with superposition coding, we can do so in the limited regimes of (R_1, R_2) , where R_1 is small enough and R_2 is large enough.

A. Encoding of Superposition Codes

In Fig. 2, we compare the codeword distributions of the random block code and of a desired UEP code with two layers of priority. Each color dot is a codeword, and the shell around it is the decoding region for $M = (M_1, M_2)$ in the output space $\{0, 1\}^N$. Here the partial message M_1 is represented by the color of the codeword. Codewords with the same color have the same partial message (MSBs) M_1 , while their M_2 's (LSBs) are different. For the random block code, the same color codewords are uniformly distributed in $\{0, 1\}^N$. When a noise vector corrupts the transmitted codeword beyond the correct decoding region, the decoded message \hat{M} may not have the same color as that of M , since there are e^{NR_1} different colors. On the other hand, if the same color codewords are concentrated together as shown in the right figure, even if noise corrupts the transmitted codeword, the color information will have higher probability of being correctly decoded. We next construct a code whose distribution satisfies such a geometric UEP property using superposition coding principles.

Superposition codes [7] were originally developed as a channel coding scheme for communications over a degraded broadcast channel. Here we use the scheme to design a UEP code for two layers of priority. A superposition code of length N and rate $R = R_1 + R_2$ is composed of $e^{N(R_1+R_2)}$ binary length- N codewords, $\{\mathbf{z}^{(m_1, m_2)}\}$, where $\mathbf{z}^{(m_1, m_2)} = \mathbf{u}^{(m_1)} \oplus \mathbf{v}^{(m_2)}$ for $m_1 \in \{0, \dots, e^{NR_1} - 1\}$, $m_2 \in \{0, \dots, e^{NR_2} - 1\}$. The partial codewords $\{\mathbf{u}^{(m_1)}\}$ and $\{\mathbf{v}^{(m_2)}\}$ are independently generated, where $u_i^{(m_1)}$'s are i.i.d. with Bernoulli(1/2) and $v_i^{(m_2)}$'s are i.i.d. with Bernoulli(α) for some $\alpha \in (0, 1/2)$.

In contrast with the random block code, where every codeword is independent, in a superposition code the subset of codewords $\{\mathbf{z}^{(m_1, m_2)}\}$, $m_2 \in \{0, \dots, e^{NR_2} - 1\}$, for a fixed m_1 (the same color codewords), are dependent on each other. Moreover, as α decreases from 1/2 to 0, codewords with the same color become more and more concentrated around $\mathbf{u}^{(m_1)}$. Therefore, the superposition code satisfies the desired geometric property of UEP for two layers of priority.

B. Decoding of Superposition Codes

Let us consider decoding of the partial messages m_1 and m_2 given channel outputs $\mathbf{y} = \mathbf{z}^{(m_1, m_2)} \oplus \mathbf{n}$ where \mathbf{n} is a

length- N noise word with i.i.d. Bernoulli(ϵ) symbols. Define the best achievable error exponents of $\Pr(\hat{M}_1 \neq M_1)$ and of $\Pr(\hat{M}_2 \neq M_2 | \hat{M}_1 = M_1)$ for the superposition code of rate (R_1, R_2) , with the parameter $\alpha \in (0, 1/2)$, as $E_{m_1}(R_1, R_2, \alpha)$ and $E_{m_2}(R_2, \alpha)$, respectively.

There have been many previous works [8], [9], [10] to analyze lower bounds on the error exponent of $\Pr(\hat{M}_1 \neq M_1)$. Even though some results were shown to achieve a numerically tighter bound than the others, there has been no closed form solution for $E_{m_1}(R_1, R_2, \alpha)$ that is exponentially tight for $\Pr(\hat{M}_1 \neq M_1)$. Among many lower bounds on $E_{m_1}(R_1, R_2, \alpha)$, we consider bounds based on two sub-optimal decoding rules, which provide meaningful results in the analysis of $\Pr(\hat{M}_1 \neq M_1)$ from the UEP perspective.

The first sub-optimal decoding rule we consider is the joint maximum likelihood decoding. Given the received word \mathbf{y} , this decoding rule finds the most probable pair of (\hat{m}_1, \hat{m}_2) . Decoding error of m_1 happens only when $\hat{m}_1 \neq m_1$ regardless of whether $\hat{m}_2 \neq m_2$ or not. When $E_{m_1, \text{JML}}(R_1, R_2)$ denotes the error exponent of $\Pr(\hat{M}_1 \neq M_1)$ with the sub-optimal joint ML decoding, it can be shown that $E_{m_1, \text{JML}}(R_1, R_2) \geq E_r(R_1 + R_2)$ for every (R_1, R_2) , regardless of the choice of $\alpha \in (0, 1/2)$. Therefore, the superposition code provides a better, or at least as good, error protection for the partial message M_1 as that of the random block code for every (R_1, R_2) , regardless of the choice of $\alpha \in (0, 1/2)$.

The second sub-optimal decoding rule we consider is successive cancellation (SC). This decoding rule behaves as if one of the partial codewords in $\{\mathbf{u}^{(m_1)}\}$ were transmitted, while treating $\mathbf{v}^{(m_2)}$ as noise. Given the channel outputs $\mathbf{y} = \mathbf{u}^{(m_1)} \oplus \mathbf{v}^{(m_2)} \oplus \mathbf{n}$, this decoder first finds the most probable $\mathbf{u}^{(\hat{m}_1)}$ under the assumption that it was transmitted through a BSC ($\alpha * \epsilon$) where $\alpha * \epsilon := \alpha(1 - \epsilon) + (1 - \alpha)\epsilon$. It then subtracts $\mathbf{u}^{(\hat{m}_1)}$ from \mathbf{y} to next decode \hat{m}_2 . Denote the error exponent of $\Pr(\hat{M}_1 \neq M_1)$ with the SC decoding as $E_{m_1, \text{SC}}(R_1, \alpha)$. In [10], it was demonstrated that

$$E_{m_1, \text{SC}}(R_1, \alpha) = \begin{cases} E_0(1/2, \alpha * \epsilon) - R_1, & 0 \leq R_1 < R_{\text{cirt}}(\alpha * \epsilon), \\ d(\gamma_{\text{GV}}(R_1) \| \alpha * \epsilon), & R_{\text{cirt}}(\alpha * \epsilon) \leq R_1 < C - C_2(\alpha), \end{cases} \quad (12)$$

where $C = h(1/2) - h(\epsilon)$, $C_2(\alpha) = h(\alpha * \epsilon) - h(\epsilon)$. It was also shown that the error exponent of $\Pr(\hat{M}_2 \neq M_2 | \hat{M}_1 = M_1)$, defined as $E_{m_2}(R_2, \alpha)$, is positive for any $R_2 \in [0, C_2(\alpha))$, while the exponent becomes 0 at $R_2 = C_2(\alpha)$.

By comparing $E_{m_1, \text{SC}}(R_1, \alpha)$ in (12) with the random block code exponent $E_r(R_1 + R_2)$ in (4), we can show that for sufficiently large R_2 and sufficiently small R_1 the superposition code provides a strictly better error protection for the partial message M_1 than the random block code does.

Moreover, for a very noisy BSC(ϵ), the following establishes a strict gain in the error exponent of $\Pr(\hat{M}_1 \neq M_1)$ for every $R_1 \in [0, C - C_2(\alpha))$ when $R_2 = C_2(\alpha)$.

Lemma 3: Assume a very noisy BSC(ϵ) with $\epsilon \approx 0.5$, a fixed $\alpha \in (0, 1/2)$, and the rate $R_2 = C_2(\alpha)$. Then the error exponent of $\Pr(\hat{M}_1 \neq M_1)$ for the superposition code, denoted

$E_{m_1}(R_1, R_2, \alpha)$, is strictly larger than that of the random block code for all $R_1 \in [0, C - C_2(\alpha)]$. More precisely,

$$E_{m_1}(R_1, R_2, \alpha) \geq E_{m_1, \text{SC}}(R_1, \alpha) > E_r(R_1 + R_2). \quad (13)$$

In Fig 3, we provide a plot of $E_r(R_1 + R_2)$ (black solid line) and $E_{m_1, \text{SC}}(R_1, \alpha)$ (blue dash-dot line) over $R = R_1 + R_2$ for a BSC(0.45) with $\alpha = 0.11$ and $R_2 = C_2(\alpha)$. It can be shown that $E_{m_1, \text{SC}}(R_1, \alpha)$ is above $E_r(R_1 + R_2)$ for every $R_1 \in [0, C - C_2(\alpha)]$. We will use the improvement in $\Pr(\hat{M}_1 \neq M_1)$ to demonstrate a gain in $\mathbb{E}[c_q(X, \hat{X}_N)]$.

C. Performance of the Superposition Code in Estimation

We demonstrate a gain in the exponent of the quantized estimation error $\mathbb{E}[c_q(X, \hat{X}_N)]$ of resolution e^{-NR} with the superposition code. Define the exponentially decreasing rate of $\mathbb{E}[c_q(X, \hat{X}_N)]$ in N for the superposition code at a fixed rate $R \in (0, C)$ as $E_s(R) = \liminf_{N \rightarrow \infty} \frac{-\ln \mathbb{E}[c_q(X, \hat{X}_N)]}{N}$. The random block code achieves an exponentially decreasing rate of $\mathbb{E}[c_q(X, \hat{X}_N)]$ equal to $E_r(R)$ in (4) for a given rate R . By using Lemma 3, we show that for a very noisy BSC, we can achieve $E_s(R) > E_r(R)$ where $R \in (\frac{C}{6}, C)$.

Theorem 1: For a very noisy BSC, the exponent $E_s(R)$ of $\mathbb{E}[c_q(X, \hat{X}_N)]$ for the superposition code is strictly larger than that of the random block code $E_r(R)$ for $R \in (\frac{C}{6}, C)$. More precisely, for any $R \in (\frac{C}{6}, C)$

$$E_s(R) \geq E_{m_1, \text{SC}}(R_1^*, \alpha^*) > E_r(R) \quad (14)$$

for α^* satisfying $C_2(\alpha^*) = \frac{6}{5}(R - \frac{C}{6})$ and $R_1^* = R - C_2(\alpha^*)$.

Sketch of Proof: Using the superposition code with two layers, the achievable $\mathbb{E}[c_q(X, \hat{X}_N)]$ can be bounded above as

$$\mathbb{E}[c_q(X, \hat{X}_N)] \leq e^{-NE_{m_1, \text{SC}}(R_1, R_2, \alpha)} + e^{-NE_{m_2}(R_2, \alpha)} e^{-2NR_1}.$$

Thus, $E_s(R) \geq \max_{(R_1, R_2, \alpha) \in B} \min\{E_{m_1, \text{SC}}, E_{m_2} + 2R_1\}$ for $B := \{(R_1, R_2, \alpha) : R_1 + R_2 = R, \alpha \in (0, 1/2)\}$. For $(R_1, R_2, \alpha) = (R - C_2(\alpha), C_2(\alpha), \alpha)$, since $E_{m_2}(R_2, \alpha) = 0$, we have $E_s(R) \geq \max_{(R_1, R_2, \alpha) \in B} \min\{E_{m_1, \text{SC}}, E_{m_2} + 2R_1\} \geq \max_{\alpha} \min\{E_{m_1, \text{SC}}, 2R_1\}$. For $R \in (\frac{C}{6}, C)$, there exists $\alpha^* \in (0, 1/2)$ that makes $E_{m_1, \text{SC}} = 2R_1 = 2(R - C_2(\alpha^*))$. Therefore, $E_s(R) \geq \min\{E_{m_1, \text{SC}}, 2R_1\} = E_{m_1, \text{SC}}(R_1, \alpha^*)$ for α^* satisfying $E_{m_1, \text{SC}}(R_1, \alpha^*) = 2(R - C_2(\alpha^*))$, or equivalently $C_2(\alpha^*) = \frac{6}{5}(R - \frac{1}{6}C)$ for $\epsilon \approx 1/2$. Moreover, as shown in Lemma 3, $E_{m_1, \text{SC}}(R_1, \alpha) > E_r(R_1 + R_2)$ for every R_1 when $R_2 = C_2(\alpha)$. Thus, for the chosen α^* and $R_1^* = R - C_2(\alpha^*)$, the inequalities in (14) hold. ■

In Fig 3, the gain in $E_s(R)$ at $R = C/2$ is illustrated.

Discussion: We next consider the achievable MSE. For the superposition code, the quantized MSE at a fixed rate R is $\mathbb{E}[c_q(X, \hat{X}_N)] \doteq e^{-NE_s(R)}$ and the MSE can be bounded as

$$\mathbb{E}[|X - \hat{X}_N|^2] \leq e^{-NE_s(R)} + e^{-N2R} \doteq e^{-N \min\{E_s(R), 2R\}}. \quad (15)$$

For the random block code, the MSE exponent is equal to $\min\{E_r(R), 2R\}$. For a very noisy BSC, we have $\min\{E_r(R), 2R\} = 2R$ for $R \in (0, C/6)$ and $\min\{E_r(R), 2R\} = E_r(R)$ for $R \in (C/6, C)$. Therefore, in the high rate regimes, the MSE exponent is dominated

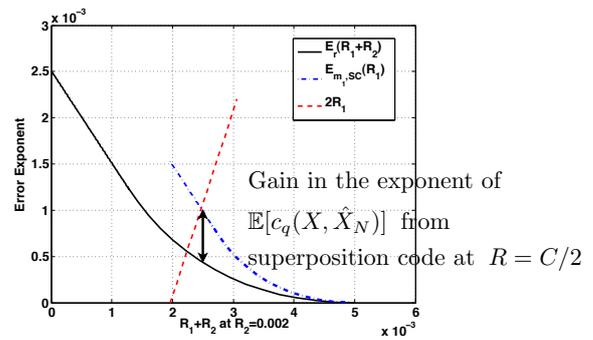


Fig. 3. A plot of $E_r(R_1 + R_2)$, $E_{m_1, \text{SC}}(R_1, \alpha)$, and $2R_1$ at $R_2 = C_2(\alpha)$ where $\epsilon = 0.45$ and $\alpha = 0.11$. When $R_2 = C_2(\alpha)$, for every $R_1 \in [0, C - C_2(\alpha)]$, $E_{m_1, \text{SC}}(R_1)$ is above $E_r(R_1 + R_2)$. For $R = C/2 = 0.0025$, by choosing α to satisfy $E_{m_1, \text{SC}}(R_1, \alpha) = 2R_1 = 2(R - C_2(\alpha))$, the achievable exponent of $\mathbb{E}[c_q(X, \hat{X}_N)]$ with the superposition code (the crossing point of $E_{m_1, \text{SC}}(R_1)$ and $2R_1$) is strictly larger than $E_r(R)$, which is the best achievable exponent of $\mathbb{E}[c_q(X, \hat{X}_N)]$ with the random block code.

by the limited reliability of decoded message. By using the superposition code, we achieved the quantized MSE exponent $E_s(R) > E_r(R)$ in $R \in (C/6, C)$. This improvement thus results in a better convergence rate of the MSE in $R \in (C/6, C)$.

IV. CONCLUSIONS

Different from classical communications, in inference problems the value of information varies significantly depending on the task-specific cost functions. We proposed an unequal error protection coding based on superposition coding for estimating a quantized random variable, with demonstrated performance gains in estimation error. A superposition code can provide unequal error protection for two layers of information bits and guarantee improved rates of convergence for the expected quantized estimation error, exponentially decreasing in N with the exponent larger than that of the random block code.

REFERENCES

- [1] E. Varnani, K. Lahouez, A. Bar-Hen, and B. Jedynek, "Non-adaptive policies for 20 questions target localization," in *Information Theory Proceedings (ISIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 775 – 778.
- [2] T. Tsiligkaridis, B. M. Sadler, and A. O. Hero, "Collaborative 20 questions for target localization," *Information Theory, IEEE Transactions on*, vol. 60, no. 4, pp. 2233–2252, 2014.
- [3] M. Horstein, "Sequential transmission using noiseless feedback," *Information Theory, IEEE Transactions on*, vol. 9, no. 3, pp. 136–143, 1963.
- [4] M. V. Burnashev and K. Zigangirov, "An interval estimation problem for controlled observations," *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 51–61, 1974.
- [5] G. D. Forney Jr, "On exponential error bounds for random codes on the bsc," unpublished manuscript, 2001.
- [6] R. G. Gallager, *Information theory and reliable communication*, Vol. 2. Wiley, 1968.
- [7] T. M. Cover, "Broadcast channels," *Information Theory, IEEE Transactions on*, vol. 18, no. 1, pp. 2–14, 1972.
- [8] Y. Kaspi and N. Merhav, "Error exponents for broadcast channels with degraded message sets," *Information Theory, IEEE Transactions on*, vol. 57, no. 1, pp. 101–123, 2011.
- [9] J. Korner and A. Sgarro, "Universally attainable error exponents for broadcast channels with degraded message sets," *Information Theory, IEEE Transactions on*, vol. 26, no. 6, pp. 670–679, 1980.
- [10] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 3–14, 1974.