Error Exponents of Optimum Decoding for the Interference Channel

Raul Etkin
Hewlett-Packard Laboratories
Palo Alto, CA 94304
Email: raul.etkin@hp.com

Neri Merhav[†]
Technion - Israel Institute of Technology
Haifa 32000, Israel
Email: merhav@ee.technion.ac.il

Erik Ordentlich
Hewlett-Packard Laboratories
Palo Alto, CA 94304
Email: erik.ordentlich@hp.com

***Abstract*— Exponential error bounds for the finite–alphabet interference channel (IFC) with two transmitter–receiver pairs, are investigated under the random coding regime. Our focus is on optimum decoding, as opposed to heuristic decoding rules that have been used in previous works, like joint typicality decoding, decoding based on interference cancellation, and decoding that considers the interference as additional noise. Indeed, the fact that the actual interfering signal is a codeword and not an i.i.d. noise process complicates the performance analysis of the optimum decoder. In addition to the single–letter expressions of the error exponents derived, we also present some numerical results and discuss them.**

I. Introduction

The M-user interference channel (IFC) models the communication between M transmitter-receiver pairs, wherein each receiver must decode its corresponding transmitter's message from a signal that is corrupted by interference from the other transmitters, in addition to channel noise. The information theoretic analysis of the IFC was initiated over 30 year ago and has recently witnessed a resurgence of interest, motivated by new potential applications, such as wireless communication over unregulated spectrum.

Previous work on the IFC has focused on obtaining inner and outer bounds to the capacity region for memoryless interference and noise, with a precise characterization of the capacity region remaining elusive for most channels, even for $M = 2$ users. The best known inner bound for the IFC is the Han-Kobayashi (HK) region, established in [1]. It has been found to be tight in certain special cases ([1], [2]), and recently was found to be tight to within 1 bit for the two user Gaussian IFC [3]. No achievable rates that lie outside the HK region are known for any IFC.

Our aim in this paper is to extend the study of achievable schemes to the analysis of error exponents, or exponential rates of decay of error probabilities, that are attainable as a function of user rates. To our knowledge, there has been no prior treatment of error exponents for the IFC. In particular, the error bounds underlying the achievability results in [1] yield vanishing error exponents (though still decaying error probability) at all rates.

[†] Part of this work was done while N. Merhav was visiting Hewlett–Packard Laboratories in the Summer of 2007.

Our main result, presented in Section II, is a single letter characterization of an achievable error exponent region, as a function of user rates, for the $M = 2$ user finite alphabet, memoryless interference channel. The region is derived by bounding the average error probability of random codebooks comprised of i.i.d. codewords uniformly distributed over a type class, under maximum likelihood (ML) decoding at each user. Unlike the single user setting, in this case, the effective channel determining each receiver's ML decoding rule is induced both by the noise *and* the interfering user's codebook. Our focus on optimal decoding is a departure from the conventional achievability arguments in [1] and elsewhere, which are based on joint-typicality decoding, with restrictions on the decoder to "treat interference as noise" or to "decode the interference" in part or in whole. However, our codebook ensembles are simpler than the superposition codebooks of [1]. It might be fruitful to consider such structured codebook ensembles from an error exponent perspective, and we plan to do so in future work [6].

The analysis of the probability of decoding error under optimal decoding is complicated due to correlations induced by the interfering signal. Usual methods for bounding the probability of error based on Jensen's inequality and other related inequalities (see, e.g., (10) in Section II) fail to give tight results. Our bounding approach combines some of the ideas of [4] and [5] used to derive error exponents for single user channels. As in [4], we use auxiliary parameters ρ and λ to get an upper bound on the average probability of decoding error under ML decoding, which we then bound using the method of types [5]. Key in our derivation is the use of distance enumerators in the spirit of [7], which allows us to avoid using Jensen's inequality in some steps, and allows us to maintain exponential tightness in other inequalities by applying them to only a polynomially few terms (as opposed to exponentially many) in certain sums that bound the probability of decoding error.

Regarding notation, unless otherwise stated, we use lowercase and uppercase letters for scalars, boldface lowercase letters for vectors, uppercase (boldface) letters for random variables (vectors), and calligraphic letters for sets. For example, a is a scalar, $\boldsymbol{v}$ is a vector, X is a random variable, $\boldsymbol{X}$ is a random vector, and $\mathcal{S}$ is a set. In addition, we write $v(t)$ to refer to the t-th element of vector $\boldsymbol{v}$. Also, we use $\log(\cdot)$ to denote natural logarithm, $\boldsymbol{E}$ to denote expectation, and Pr to denote probability. For independent random variables X and Y distributed according to $P_{X,Y}(x,y) = P_X(x)P_Y(y)$, $(x,y) \in \mathcal{X} \times \mathcal{Y}$, we define the operator $\boldsymbol{E}_X(\cdot)$ as $\boldsymbol{E}_X(f(X,Y)) \triangleq \sum_{x\in\mathcal{X}} f(x,Y)P_X(x)$ for any function $f(\cdot,\cdot)$. All information quantities (entropy, mutual information, etc.) and rates are in nats. Finally, we use $\doteq$, $\stackrel{.}{\leq}$, etc., to denote equality or inequality to the first order in the exponent, i.e. $a_n \doteq b_n \Leftrightarrow \lim_{n\to\infty} \frac{1}{n}\log\frac{a_n}{b_n} = 0$; $a_n \stackrel{.}{\leq} b_n \Leftrightarrow \limsup \frac{1}{n}\log\frac{a_n}{b_n} \leq 0$.

We continue with a formal description of the two user IFC setting. Let $\boldsymbol{x}_i = (x_i(1),\ldots,x_i(n)) \in \mathcal{X}_i^n$, $i = 1,2$, denote the channel input signals of the two transmitters, and let $\boldsymbol{y}_i = (y_i(1),\ldots,y_i(n)) \in \mathcal{Y}_i^n$ be the corresponding channel outputs received by decoders 1 and 2, where $\mathcal{X}_i$ and $\mathcal{Y}_i$ denote the input and output alphabets, and which we assume to be finite. Each (random) output symbol pair $(Y_1(j), Y_2(j))$ is assumed to be conditionally independent of all other outputs, and all input symbols, given the two corresponding (random) input symbols $(X_1(j), X_2(j))$, and the corresponding conditional probability is assumed to be constant from symbol to symbol. An (n, R_1, R_2) code for the IFC consists of pairs of encoding and decoding functions, (f_1, f_2) and (g_1, g_2), respectively, where $f_i : \{1,\ldots,\lceil e^{nR_i}\rceil\} \to \mathcal{X}_i^n$ and $g_i : \mathcal{Y}_i^n \to \{1,\ldots,\lceil e^{nR_i}\rceil\}$. The performance of the code is characterized by a pair of error probabilities $P_{e,i} = \Pr(\hat{W}_i \neq W_i)$, $i = 1,2$, where $\hat{W}_i = g_i(\boldsymbol{Y}_i)$ and $\boldsymbol{Y}_i$ is the random output when user i transmits $\boldsymbol{X}_i = f_i(W_i)$, assuming the messages W_i are uniformly distributed on the sets of indices $\{1,2,\ldots,\lceil e^{nR_i}\rceil\}$, $i = 1,2$. The per user error probabilities depend on the channel only through the marginal conditional distributions of the channel outputs given the corresponding channel input pairs. We shall denote these conditional distributions as $q_i(y|x_1,x_2) \triangleq \Pr(Y_i(j) = y|(X_1(j), X_2(j)) = (x_1,x_2))$.

A pair of error exponents (E_1, E_2) is attainable at a rate pair (R_1, R_2) if there is a sequence of (n, R_1, R_2) codes satisfying $E_i \leq \liminf -(1/n)\log P_{e,i}$ for $i = 1,2$. The set of all attainable error exponents at (R_1, R_2) comprises the error exponent region at (R_1, R_2) and we shall denote it as $\mathcal{E}(R_1, R_2)$. The main result of this paper is a single letter characterization of a non-trivial subset of $\mathcal{E}(R_1, R_2)$ for each R_1, R_2.

Before presenting the main result, we first derive an "easy" set of attainable error exponents which we shall treat as a benchmark for the more sophisticated exponents of the next section. The "easy" exponents are obtained from Gallager's single user random coding error exponents for suitable "average" channels.

Given distributions Q_i on $\mathcal{X}_i$, let $\overline{q}_i(y|x)$ denote the average channel induced for user i if user j's transmitted symbol, $j \neq i$, is distributed according to Q_j. That is, $\overline{q}_1(y_1|x_1) = \sum_{x_2\in\mathcal{X}_2} q_1(y_1|x_1,x_2)Q_2(x_2)$, with $\overline{q}_2(y_2|x_2)$ defined analogously. It is reasonable to expect that, for $i = 1,2$, Gallager's random coding error exponents corresponding to input distributions Q_i and induced single user channels $\overline{q}_i$ are attainable. From eqs. (5.6.13) and (5.6.14) in [4], for $i = 1,2$, these exponents correspond to

$$E_{G,i} = \max_{0\leq\rho\leq 1}\Bigg\{ -\rho R_i - \log\sum_{y\in\mathcal{Y}_i}\Bigg[\sum_{x\in\mathcal{X}_i} Q_i(x)\overline{q}_i(y|x)^{1/(1+\rho)}\Bigg]^{1+\rho}\Bigg\}. \quad (1)$$

The following simple argument shows that these exponents are indeed achievable. Suppose each receiver implements an ML decoder assuming a discrete memoryless channel (DMC) with transition probabilities given by its corresponding average channel $\overline{q}_i$. Unlike what will be treated in the next section, these are suboptimal decoders, since the true induced channels depend on the interfering users' codebooks. The error probabilities corresponding to these simpler decoders ($i = 1,2$) can be written as

$$P_{e,i} = \frac{1}{\lceil e^{nR_1}\rceil\lceil e^{nR_2}\rceil}\sum_{\boldsymbol{x}_1\in C_1}\sum_{\boldsymbol{x}_2\in C_2}\sum_{\boldsymbol{y}\in\mathcal{Y}_i^n} q_i^{(n)}(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2)\cdot 1(\exists \boldsymbol{x}' \neq \boldsymbol{x}_1, \boldsymbol{x}' \in C_1 : \overline{q}_i^{(n)}(\boldsymbol{y}|\boldsymbol{x}_1) \leq \overline{q}_i^{(n)}(\boldsymbol{y}|\boldsymbol{x}')) \quad (2)$$

where C_i is the codebook of user i, $q_i^{(n)}(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2) = \prod_{t=1}^n q_i(y(t)|x_1(t),x_2(t))$, $\overline{q}_i^{(n)}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{t=1}^n \overline{q}_i(y(t)|x(t))$, and $1(\cdot)$ denotes the indicator function. Assuming the symbols across all codewords in C_i are selected i.i.d. according to the product distribution Q_i, the expectation of $P_{e,1}$ over the random codebooks C_1 and C_2, denoted as $\overline{P}_{e,1}$, is given by

$$\overline{P}_{e,1} = \boldsymbol{E}_{C_1}\Bigg[\frac{1}{\lceil e^{nR_1}\rceil}\sum_{\boldsymbol{x}_1\in C_1}\sum_{\boldsymbol{y}\in\mathcal{Y}_i^n}\overline{q}_i^{(n)}(\boldsymbol{y}|\boldsymbol{x}_1)\cdot 1(\exists \boldsymbol{x}' \neq \boldsymbol{x}_1, \boldsymbol{x}' \in C_1 : \overline{q}_1^{(n)}(\boldsymbol{y}|\boldsymbol{x}_1) \leq \overline{q}_1^{(n)}(\boldsymbol{y}|\boldsymbol{x}'))\Bigg], \quad (3)$$

with a similar expression holding for $\overline{P}_{e,2}$. In particular, only the terms $q_i^{(n)}(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2)$ in (2) depend on C_2, and averaging them over C_2 (selected according to the product distribution) yields the terms $\overline{q}_i^{(n)}(\boldsymbol{y}|\boldsymbol{x}_1)$. The expression (3), however, corresponds exactly to the expected error probability (with respect to the random codebook C_1) of single user ML decoding for the "averaged" DMC $\overline{q}_1$, and the exponential behavior of this, as is well known from [4], is indeed bounded from below by $E_{G,1}$ of (1). This (and the analogous argument for $E_{G,2}$) establishes that $E_{G,1}$ and $E_{G,2}$ are indeed attainable exponents for the IFC.

In the next section, we derive a more sophisticated set of attainable exponents by analyzing true ML decoding for the channel induced by the interfering codebook. We follow this up in Section III with a numerical comparison of the new exponents with $E_{G,1}$ and $E_{G,2}$ for a simple IFC. These results show that our improved exponents are never worse, and, for most rates, strictly improve over $E_{G,1}$ and $E_{G,2}$.

II. Main Result

Our main contribution is stated in the following theorem, which presents a new error exponent region for the discrete memoryless two-user IFC.

Theorem 1: For a discrete memoryless two-user IFC as defined in Section I, for a family of block codes of rates R_1 and R_2 a decoding error probability for user 1 satisfying

$$\liminf -\frac{1}{n}\log \bar{P}_{e,1}(n) \geq E_{R,1}(R_1, R_2, Q_1, Q_2, \rho, \lambda) \quad (4)$$

can be achieved as the block length of the codes n goes to infinity, where the error exponent $E_{R,1}(R_1, R_2, Q_1, Q_2, \rho, \lambda)$ is given by

$$E_{R,1} = \Bigg\{ R_2 - \rho R_1 + \min\Bigg\{ \min_{\substack{(P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}, P_{\hat{X}'_1,\hat{X}'_2,\hat{Y}'_1}) \\ \in \mathcal{S}_1(Q_1,Q_2)}} f_1\left(\rho,\lambda, P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}, P_{\hat{X}'_1,\hat{X}'_2,\hat{Y}'_1}\right); \min_{\substack{(P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}, P_{\hat{X}'_1,\hat{X}'_2,\hat{Y}'_1}) \\ \in \mathcal{S}_2(Q_1,Q_2,R_2)}} f_2\left(\rho,\lambda, P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}, P_{\hat{X}'_1,\hat{X}'_2,\hat{Y}'_1}\right) \Bigg\}\Bigg\} \quad (5)$$

where

$$\begin{aligned} f_1 \triangleq\, & g(\rho,\lambda, P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}, P_{\hat{X}'_1,\hat{X}'_2,\hat{Y}'_1}) - H(\hat{Y}_1|\hat{X}_1) + \rho I(\hat{X}'_1;\hat{Y}'_1) \\ & + \max\Big\{ I(\hat{X}_2;\hat{X}_1,\hat{Y}_1) - R_2; \\ & \qquad (1-\rho\lambda)(I(\hat{X}_2;\hat{X}_1,\hat{Y}_1) - R_2)\Big\} \\ & + \max\Big\{ (1-\rho) I(\hat{X}'_2;\hat{Y}'_1) + \rho I(\hat{X}'_2;\hat{X}'_1,\hat{Y}'_1) - R_2; \\ & \rho(I(\hat{X}'_2;\hat{X}'_1,\hat{Y}'_1) - R_2); \rho\lambda(I(\hat{X}'_2;\hat{X}'_1,\hat{Y}'_1) - R_2)\Big\} \end{aligned} \quad (6)$$

$$\begin{aligned} f_2 \triangleq\, & g(\rho,\lambda, P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}, P_{\hat{X}'_1,\hat{X}'_2,\hat{Y}'_1}) - H(\hat{Y}_1|\hat{X}_1) \\ & + \rho I(\hat{X}'_1;\hat{X}'_2,\hat{Y}'_1) + I(\hat{X}_2;\hat{X}_1,\hat{Y}_1) - R_2 \end{aligned} \quad (7)$$

with

$$\begin{aligned} g \triangleq & -(1-\rho\lambda) E_{\hat{X}_1,\hat{X}_2,\hat{Y}_1} \log q_1(\hat{Y}_1|\hat{X}_1,\hat{X}_2) \\ & - \rho\lambda E_{\hat{X}'_1,\hat{X}'_2,\hat{Y}'_1} \log q_1(\hat{Y}'_1|\hat{X}'_1,\hat{X}'_2) \end{aligned}$$

and

$$\begin{aligned} \mathcal{S}_1(Q_1,Q_2) \triangleq & \{(P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}, P_{\hat{X}'_1,\hat{X}'_2,\hat{Y}'_1}) \in \mathcal{S}^2 : P_{\hat{Y}_1} = P_{\hat{Y}'_1}, \\ & P_{\hat{X}_1} = P_{\hat{X}'_1} = Q_1, P_{\hat{X}_2} = P_{\hat{X}'_2} = Q_2\} \end{aligned} \quad (8)$$

$$\begin{aligned} \mathcal{S}_2(Q_1,Q_2,R_2) \triangleq & \{(P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}, P_{\hat{X}'_1,\hat{X}'_2,\hat{Y}'_1}) \in \mathcal{S}^2 : \\ & P_{\hat{X}_1} = P_{\hat{X}'_1} = Q_1, P_{\hat{X}_2} = P_{\hat{X}'_2} = Q_2, \\ & R_2 \leq I(\hat{X}_2;\hat{Y}_1), P_{\hat{X}_2,\hat{Y}_1} = P_{\hat{X}'_2,\hat{Y}'_1}\} \end{aligned} \quad (9)$$

where $\mathcal{S}$ is the probability simplex in $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}_1$. In the bound (4), $(\rho,\lambda) \in [0,1]^2$ can be chosen to maximize the error exponent $E_{R,1}$.

In eqs. (4), (5), (8), and (9), Q_1 and Q_2 are probability distributions defined over the alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively.

Expressions for the error probability $P_{e,2}$ and error exponent $E_{R,2}$ equivalent to (4) and (5) can be stated for the receiver of user 2 by replacing $X_1 \leftrightarrow X_2$, $Y_1 \rightarrow Y_2$, and $q_1 \rightarrow q_2$ in all the expressions. By varying Q_1 and Q_2 over all probability distributions in $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively, we obtain the error exponent region for fixed rates R_1 and R_2.

Remark: The set of rate pairs (R_1, R_2) for which the corresponding error exponent regions contain points with strictly positive components can be shown to be contained in the HK region [8]. A precise characterization of this set of rate pairs is left for future work.

Proof Outline: Due to space limitations we will omit some of the details of the derivation. The complete proof will be presented in [6]. We will use the following inequality (see problem 4.15, part (f) in [4]), valid for $a_i \geq 0$, $i = 1,\ldots,n$ and $0 \leq b \leq 1$:

$$\Big(\sum_{i=1}^{n} a_i\Big)^b \leq \sum_{i=1}^{n} a_i^b. \quad (10)$$

For a given block length n, we generate the codebook of user $i = 1,2$ by choosing $M_i \triangleq \lceil e^{nR_i} \rceil$ sequences $\boldsymbol{x}_i$ of length n independently and uniformly over all the sequences of length n and type Q_i in $\mathcal{X}_i^n$. We will write $\boldsymbol{x}_{i,j}$ to denote the j-th codeword of user i. For the moment, we make the technical assumption that $Q_i, i = 1,2$ have rational entries with denominator n.

For a given channel output $\boldsymbol{y}_1 \in \mathcal{Y}_1^n$, the best decoding rule to minimize the probability of error in decoding the message of user 1 is ML decoding, which consists of picking the message m which maximizes $P(\boldsymbol{y}_1|\boldsymbol{x}_{1,m}) = \sum_{i=1}^{M_2} q_1^{(n)}(\boldsymbol{y}_1|\boldsymbol{x}_{1,m},\boldsymbol{x}_{2,i})/M_2$. Letting[1]

$$q_{1,C_2}^{(n)}(\boldsymbol{y}_1|\boldsymbol{x}_1) \triangleq \frac{1}{M_2}\sum_{i=1}^{M_2} q_1^{(n)}(\boldsymbol{y}_1|\boldsymbol{x}_1,\boldsymbol{x}_{2,i}) \quad (11)$$

be the "average" channel observed at receiver 1, where the averaging is done over the codewords of user 2 in C_2, the decoding error probability at receiver 1 for transmitted codeword $\boldsymbol{x}_{1,m}$ and codebooks C_1 and C_2 is given by:

$$P_{e,1}(\boldsymbol{x}_{1,m}, C_1, C_2) = \sum_{\boldsymbol{y}_1 \in \mathcal{Y}_1^n} P_{e,1}(\boldsymbol{x}_{1,m}, C_1, C_2|\boldsymbol{y}_1) q_{1,C_2}^{(n)}(\boldsymbol{y}_1|\boldsymbol{x}_{1,m}) \quad (12)$$

With the introduction of the average channel (11), and the use of two auxiliary parameters $(\rho,\lambda) \in [0,1]^2$, we can follow the approach of [4] to bound the conditional probability of decoding error $P_{e,1}(\boldsymbol{x}_m, C_1, C_2|\boldsymbol{y}_1)$. Taking expectation over the random choice of codebooks C_1 and C_2 we obtain an average error probability:

$$\begin{aligned} \bar{P}_{E_1} \leq & M_1^{\rho} \sum_{\boldsymbol{y}_1 \in \mathcal{Y}_1^n} \boldsymbol{E}_{C_2}\Big\{ \boldsymbol{E}_{\boldsymbol{X}_1}\Big[[q_{1,C_2}^{(n)}(\boldsymbol{y}_1|\boldsymbol{X}_1)]^{1-\rho\lambda} \Big] \\ & \cdot \boldsymbol{E}^{\rho}_{\boldsymbol{X}_1}\Big[[q_{1,C_2}^{(n)}(\boldsymbol{y}_1|\boldsymbol{X}_1)]^{\lambda} \Big] \Big\} \end{aligned} \quad (13)$$

where we used Jensen's inequality in the last step.

Equation (13) is hard to handle, mainly due to the correlation introduced by C_2 between the two factors inside the

[1] Note that this average channel differs from the one used in Section I due to the difference in the codebook generation process.

outer expectation. Furthermore, the evaluation of the inner expectations over $\boldsymbol{X}_1$ are complicated due to the powers $(1-\rho\lambda)$ and λ affecting $q^{(n)}_{1,C_2}(\boldsymbol{y}_1|\boldsymbol{X}_1)$. Bounding methods based on using Jensen's inequality and (10) fail to give good results due to the loss of exponential tightness.

We proceed with a refined bounding technique based on the method of types inspired by [7]. While in this approach we still use (10), we use it to bound sums with a number of terms that only grows polynomially with n, and as a result, exponential tightness is preserved.

Since the channel is memoryless,

$$\begin{aligned} q^{(n)}_{1,C_2}(\boldsymbol{y}_1|\boldsymbol{x}_1) =& \frac{1}{M_2}\sum_{i=1}^{M_2}\prod_{t=1}^{n} q_1(y_1(t)|x_1(t),x_{2,i}(t)) \\ =& \frac{1}{M_2}\sum_{P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}} N_{\boldsymbol{x}_1,\boldsymbol{y}_1,C_2}(P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}) \\ & \cdot e^{nE_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}\left[\log q_1(\hat{Y}_1|\hat{X}_1,\hat{X}_2)\right]} \end{aligned} \tag{14}$$

where we used $N_{\boldsymbol{x}_{1,m},\boldsymbol{y}_1,C_2}(P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1})$ to denote the number of codewords $\boldsymbol{x}_2$ in C_2 such that $(\boldsymbol{x}_{1,m},\boldsymbol{x}_2,\boldsymbol{y}_1)$ have empirical distribution $P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}$. We also used $E_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}(\cdot)$ to denote expectation with respect to the distribution $P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}$.

Replacing (14) in (13) and using (10) three times we obtain:

$$\begin{aligned} \bar{P}_{E_1} \leq& \frac{M_1^{\rho}}{M_2}\sum_{\hat{P}}\sum_{\hat{P}'}\sum_{\boldsymbol{y}_1\in\mathcal{Y}_1^n} \boldsymbol{E}_{C_2}\Bigg\{\boldsymbol{E}_{\boldsymbol{X}_1}\left[N^{1-\rho\lambda}_{\boldsymbol{X}_1,\boldsymbol{y}_1,C_2}(\hat{P})\right] \\ & \cdot \boldsymbol{E}^{\rho}_{\boldsymbol{X}_1}\left[N^{\lambda}_{\boldsymbol{X}_1,\boldsymbol{y}_1,C_2}(\hat{P}')\right]\Bigg\} \\ & \cdot e^{n[(1-\rho\lambda)E_{\hat{P}}\log q_1(\hat{Y}_1|\hat{X}_1,\hat{X}_2)+\lambda E_{\hat{P}'}\log q_1(\hat{Y}_1'|\hat{X}_1',\hat{X}_2')} \end{aligned} \tag{15}$$

where we used $\hat{P}=P_{\hat{X}_1,\hat{X}_2,\hat{Y}_1}$ and $\hat{P}'=P_{\hat{X}_1',\hat{X}_2',\hat{Y}_1'}$ to shorten the expression.

We next consider the bounding of

$$\begin{aligned} & E(\boldsymbol{y}_1,\hat{P},\hat{P}') \triangleq \\ & \boldsymbol{E}_{C_2}\left\{\boldsymbol{E}_{\boldsymbol{X}_1}\left[N^{1-\rho\lambda}_{\boldsymbol{X}_1,\boldsymbol{y}_1,C_2}(\hat{P})\right]\boldsymbol{E}^{\rho}_{\boldsymbol{X}_1}\left[N^{\lambda}_{\boldsymbol{X}_1,\boldsymbol{y}_1,C_2}(\hat{P}')\right]\right\}, \end{aligned} \tag{16}$$

and note that $N_{\boldsymbol{X}_1,\boldsymbol{y}_1,C_2}(\hat{P})$ and $N_{\boldsymbol{X}_1,\boldsymbol{y}_1,C_2}(\hat{P}')$ are formed by sums of an exponentially large number of indicator functions, each of which takes value 1 with exponentially small probability. These sums concentrate around their means, which show different behavior depending on how the number of terms in the sum (e^{nR_2}) compares to the probability of each of the indicator functions taking value 1 (depending on the case considered, these probabilities take the form $e^{-nI(\hat{X}_2;\hat{X}_1,\hat{Y}_1)}$, $e^{-nI(\hat{X}_2';\hat{X}_1',\hat{Y}_1')}$, or $e^{-nI(\hat{X}_2';\hat{Y}_1')}$). Whenever one of the factors in (16) concentrates around its mean it behaves as a constant, and hence is uncorrelated with the remaining factor. As a result, the correlation between the two factors of (16), which complicates the analysis, can be circumvented. We omit the details of this part of the derivation, but note that the resulting bound on $E(\boldsymbol{y}_1,\hat{P},\hat{P}')$ depends on $\boldsymbol{y}_1$ only through a factor $1(\boldsymbol{y}_1\in P_{\hat{Y}_1},P_{\hat{Y}_1'};P_{\hat{X}_1}=P_{\hat{X}_1'}=Q_1;P_{\hat{X}_2}=P_{\hat{X}_2'}=Q_2)$. Therefore, the innermost sum in (15) can be evaluated by counting the number of vectors $\boldsymbol{y}_1\in\mathcal{Y}_1^n$ that have empirical types $P_{\hat{Y}_1}$ and $P_{\hat{Y}_1'}$. Note that this count can only be positive for $P_{\hat{Y}_1}=P_{\hat{Y}_1'}$. This count is approximately equal to $e^{nH(\hat{Y}_1)}$ to first order in the exponent. Furthermore, the sums over $\hat{P}$ and $\hat{P}'$ in (15) have a number of terms that only grows polynomially with n. Therefore, to first order, the exponent of (15) equals the maximum exponent of the argument of the outer two sums, where the maximization is performed over the distributions $\hat{P}$ and $\hat{P}'$ which are rational, with denominator n. We can further upper bound the probability of error by enlarging the optimization region, maximizing over *any* probability distributions $\hat{P},\hat{P}'$.

So far, we have assumed rational distributions Q_1,Q_2, and showed that (4) can be achieved. It is possible to show that $E_{R,1}(R_1,R_2,Q_1,Q_2,\rho,\lambda))$ (cf. (5)) is a continuous function of Q_1 and Q_2. It follows that for fixed ρ and λ the error exponent obtained with *any* Q_1 and Q_2 can be asymptotically achieved by using a sequence of rational $\{Q_{1,n},Q_{2,n}\}_n$ which converges to Q_1, Q_2 as $n\to\infty$. Finally, ρ and λ can be optimized to maximize the resulting error exponent. ■

III. Numerical Results

In this section we present a numerical example to show the performance of the error exponent region introduced in Theorem 1. We use as a baseline for comparison the error exponent region of Section I which is an extension of Gallager's results for single user channels to the IFC.

We present preliminary results for the binary Z-channel model: $Y_1=X_1*X_2\oplus Z_1$, $Y_2=X_2$, where $X_1,X_2,Y_1,Y_2\in\{0,1\}$, $Z_1\sim$ Bernoulli(p), $*$ is multiplication, and $\oplus$ is modulo 2 addition. This is a modified version of the binary erasure IFC that we studied in [9], where we added noise Z_1 to the received signal of user 1. In the results presented here, we fixed $p=0.01$.

The error exponent region is a surface in four dimensions $R_1,R_2,E_{R,1},E_{R,2}$. In order to obtain two-dimensional plots we consider two projections:

- Fix R_2 and maximize $E_{R,1}$ subject to $E_{R,2}>0$, varying R_1 (cf. figs. 1 and 2).
- Fix R_2 and maximize $\min\{E_{R,1},E_{R,2}\}$, varying R_1 (cf. fig. 3).

In the first projection, we study the maximum error exponent possible for user 1, only requiring reliable communication for user 2. In the second projection we study the maximum error exponent *simultaneously* achievable for both users.

Fig. 1 shows that the curves of $E_{R,1}$ for fixed Q_2 have a linear part for R_1 below a critical value, and a curvy part for R_1 above this value. This behavior is also observed in the single user random coding exponent of [4], and as a result, it also appears in the curves of $E_{G,1}$.

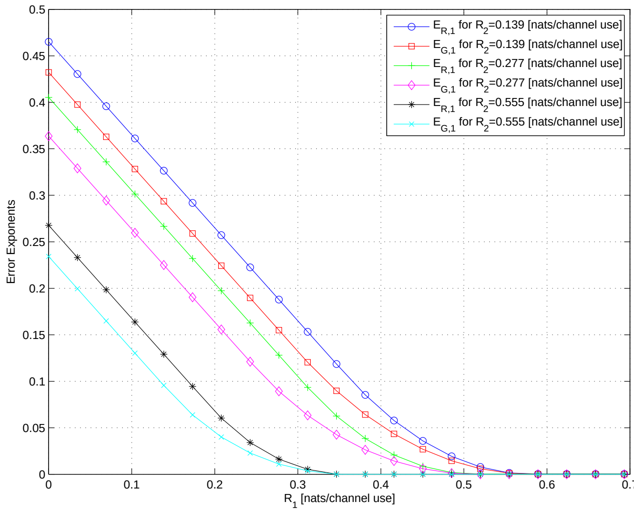

Fig. 1. Error exponent of user 1 for fixed R_2 as a function of R_1. The random codebook distribution for user 2 is optimized to maximize the error exponent of user 1 while achieving R_2.

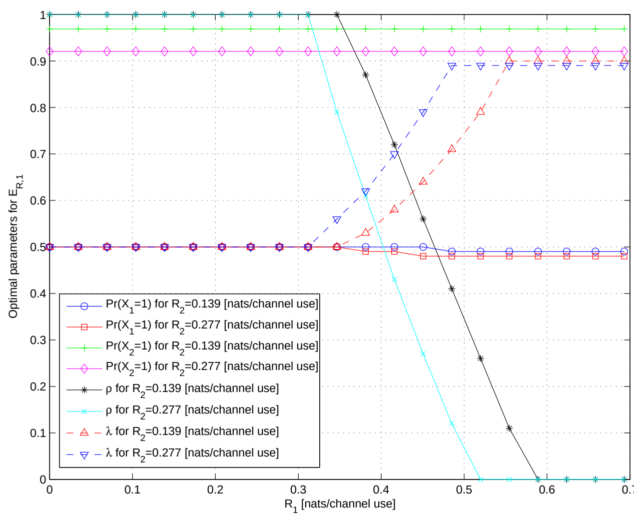

Fig. 2. Optimal parameters that maximize $E_{R,1}$ for fixed R_2 as a function of R_1. The random codebook distribution for user 2 is optimized to maximize the error exponent of user 1 while achieving R_2.

Fig. 2 shows the optimal parameters for the $E_{R,1}$ curves shown in fig. 1 for $R_2 = 0.139$ and $R_2 = 0.277$ nats/channel use. We note that since the input alphabets are binary, Q_i is completely determined by $\Pr(X_i = 1)$. Since Q_2 is chosen so that $\Pr(X_2 = 1) \geq 1/2$ and $H(X_2) = R_2$, $\Pr(X_2 = 1)$ does not vary with R_1 and decreases toward $1/2$ for increasing R_2. We see from fig. 2 that for small values of R_1, $\rho = 1$ is optimal, while for larger values of R_1, the optimal ρ decreases gradually to 0. On the other hand, for small values of R_1, $\lambda = 1/2$ is optimal, while for larger values of R_1, the optimal λ increases gradually toward 1.

Fig. 3 is obtained by choosing Q_2 to maximize $\min\{E_{R,1}, E_{R,2}\}$. For the noiseless binary channel of user 2, $E_{R,2} = \max\{H(Q_2) - R_2; 0\}$, and as a result, $E_{R,2}$ decreases with increasing $\Pr(X_2 = 1)$ for $\Pr(X_2 = 1) \geq 1/2$. On the other hand, because of the multiplication between X_1 and X_2 in the received signal Y_1, increasing $\Pr(X_2 = 1)$ results in

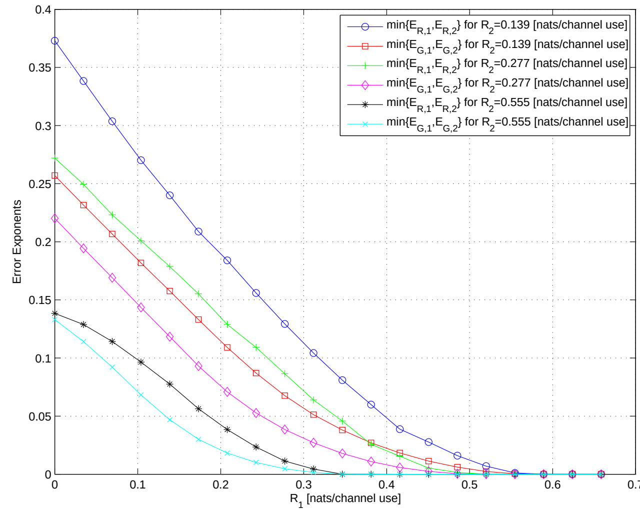

Fig. 3. Maximum error exponent simultaneously achievable for both users for fixed R_2 as a function of R_1.

less interference for user 1, and a larger value of $E_{R,1}$. It follows that there is a direct trade-off between $E_{R,1}$ and $E_{R,2}$ through the choice of Q_2, and whenever $\min\{E_{R,1}, E_{R,2}\}$ is maximized, $E_{R,1} = E_{R,2}$. Therefore, in the curves of fig. 3, $E_{R,1} = E_{R,2}$.

From the plots of figs. 1 and 3 we see that the error exponents obtained from Theorem 1 always outperform the baseline error exponents of Section I. It is worthwhile to note that the random codebook distributions used to compute $E_{R,i}$ and $E_{G,i}$ are not the same. $E_{R,i}$ is obtained using codebooks generated by choosing the codewords uniformly and independently over all sequences of length n and type Q_i. On the other hand, $E_{G,i}$ is computed using codebooks generated by choosing the codewords with n i.i.d. symbols drawn from Q_i. The performance improvement of $E_{R,i}$ over $E_{G,i}$ can be attributed to both the different random codebook distributions and the improved decoding rule (ML vs. suboptimal decoding).

References

[1] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Info. Theory*, vol. IT-27, pp. 49 - 60, January 1981.

[2] A. B. Carleial, "A case where interference does not reduce capacity," *(Corresp.), IEEE Trans. Info. Theory*, vol. IT-21, pp. 569 - 570, September 1975.

[3] R. Etkin, D. Tse, and H. Wang, "Gaussian Interference Channel Capacity to Within One Bit," *submitted to IEEE Transactions on Information Theory*, Feb. 2007. Also, available on–line at: [http://arxiv.org/PS_cache/cs/pdf/0702/0702045v2.pdf].

[4] R. G. Gallager, "Information Theory and Reliable Communication," *John Wiley & Sons, Inc. New York, NY, USA*, 1968.

[5] I. Csiszár, and J. Körner, "Information Theory: Coding Theorems for Discrete Memoryless Systems," Akadémiai Kiadó, Budapest, 1981.

[6] R. Etkin, N. Merhav, E. Ordentlich, "Error Exponents of Optimum Decoding for the Interference Channel," *in preparation*.

[7] N. Merhav, "Relations between random coding exponents and the statistical physics of random codes," *submitted to IEEE Trans. Inform. Theory*, August 2007. Also, available on–line at: [http://www.ee.technion.ac.il/people/merhav/papers/p117.pdf].

[8] C. Chang, Private communication, March 2008.

[9] R. Etkin, E. Ordentlich, "Discrete Memoryless Interference Channel: New Outer Bound," *Proceedings of the IEEE International Symposium on Information Theory*, Nice, France, June 2007.