On a Class of Time-Varying Gaussian ISI Channels

Kamyar Moshksar

Abstract— This paper studies a class of stochastic and time-varying Gaussian intersymbol interference (ISI) channels. The probability law for the $i^{th}$ channel tap during time slot $t$ is supported over an interval of centre $c_i$ and radius $r_i$. The transmitter and the receiver only know the centres $c_i$ and the radii $r_i$. The joint distribution for the array of channel taps and their realizations are unknown to both the transmitter and the receiver. A lower bound (achievability result) is presented on the channel capacity which results in an upper bound on the capacity loss compared to when all radii are zeros. The lower bound on the channel capacity saturates at a positive value as the maximum average input power $P$ increases beyond what is referred to as the saturation power $P_{sat}$. Roughly speaking, $P_{sat}$ is inversely proportional to the sum of the squares of the radii $r_i$. It is also verified that in the presence of channel state information at the receiver, if the so-called central channel frequency response is everywhere nonzero, then the aforementioned capacity loss is bounded from above by a constant that does not depend on $P$.

A partial converse result is provided in a scenario where different channel taps vary independently of each other and each channel tap process is stationary with a finite differential entropy rate. It is shown that for every sequence of codebooks with vanishing probability of error, if the size of each symbol in every codeword is bounded away from zero by a constant that is proportional to $\sqrt{P}$, then the rate of that sequence of codebooks does not scale with $P$. Tools in matrix analysis such as matrix norms and Weyl’s inequality on perturbation of eigenvalues of symmetric matrices are used in order to analyze the probability of error. A result in the paper that may find other applications is a new and tight upper bound on the size of a Gaussian typical set.

Index Terms— Intersymbol interference, time-varying channels, joint-typicality decoding, matrix norm, Weyl’s inequality.

I. INTRODUCTION

A. Summary of Prior Art

Transmission beyond the Nyquist rate over a band-limited communication channel results in a phenomenon known as intersymbol interference (ISI). The time-invariant Gaussian ISI channel is modelled by a linear filter with known finite impulse-response and additive white Gaussian noise [1]. The capacity for this channel was first studied in [19] where it was shown that Gaussian signalling is capacity-achieving. Addressing more practical structures for signal transmission, the Gaussian ISI channel with independent and identically distributed (I.I.D.) signalling over a fixed finite alphabet was examined in [3] and more recently in [4] and [5].

B. Summary of Contributions

Consider a Gaussian ISI channel initially modelled by a sequence of filter taps $c_0, c_1, \cdots, c_k$ as its impulse response. Let $h_{t,i}$ be the actual $i^{th}$ channel tap during time slot $t$. Then the initial model assumes $h_{t,i} = c_i$ for all $t$ and $i$. We ask the following question:

\text{If } h_{t,i} \text{ for } t = 1, 2, \cdots \text{ are not exactly the constant } c_i, \text{ but a random process with unknown dynamics that takes values near } c_i, \text{ then how would that impact the channel capacity?}

In an attempt to answer this question, it is assumed that the probability law for $h_{t,i}$ is supported over the interval $[c_i - r_i, c_i + r_i]$ for all $t$ and $i$ where the centres $c_i$ and the radii $r_i \geq 0$ are known constants to the transmitter.
and the receiver. The joint distribution for the array $h_{t,i}$ (including its marginals) as well as the realizations of $h_{t,i}$ remain unknown to both the transmitter and the receiver. The capacity $C^*$ for this channel is studied subject to a maximum average transmission power $P$. We consider mismatched Gaussian joint-typicality decoding tuned to the channel taps $c_0, c_1, \ldots, c_k$ and investigate the performance of the ensemble of Gaussian codebooks where the codewords are independent zero-mean Gaussian vectors with common covariance matrix $\Sigma_n$. Here, $n$ denotes the length of the codewords. It is shown that if the sequence of positive-definite matrices $\Sigma_n$ satisfies the condition $\lim_{n \to \infty} e^{-\frac{1}{n} \lambda_{\max} (\Sigma_n)} = 0$ where $\lambda_{\max} (\Sigma_n)$ is the largest eigenvalue of $\Sigma_n$, then the probability of decoding error is guaranteed to vanish as $n$ grows to infinity for sufficiently small values of the transmission rate. As a result, we obtain a lower bound $C^*_L$ on the capacity $C^*$ which is presented in Theorem 1 in Section II. This lower bound is further maximized in Proposition 1 in Section III over the admissible set of matrices $\Sigma_n$ where water-filling is performed. A feature of $C^*_L$ is that it saturates at a positive level as $P$ increases beyond what we refer to as the saturation power denoted by $P_{sat}$. Its value is given by $P_{sat} = \frac{2}{(k+1)\sum c_i^2} - 2J$ where $J$ depends entirely on the coefficients $c_i$. An upper bound is established in Corollary 2 in Section III on the amount of loss in channel capacity before saturation occurs compared to when all radii $r_i$ are zeros. We also present an upper bound on the capacity loss when channel state information is available at the receiver end. It is shown that if the so-called central channel frequency response is everywhere nonzero, then one can achieve rates that are within a constant gap (a gap that does not depend on $P$) to the capacity in the absence of channel variations. Tools in matrix analysis such as matrix norms and Weyl’s inequality on perturbation of eigenvalues of symmetric matrices are used in order to analyze the probability of error.

A partial converse result is presented in Theorem 2 in Section II. It addresses the scenario where the array $h_{t,i}$ varies independently along the index $i$, i.e., different channel taps evolve independently of each other and it is stationary along the index $t$ with a finite differential entropy rate. It is observed that for a sequence of codebooks with vanishing probability of error, if there exists a constant $a > 0$ such that the size of every symbol in each codeword is at least $a\sqrt{P}$, then the rate of that sequence of codebooks does not scale with $P$. Furthermore, it is argued that by eliminating ISI through zero-padding at the transmitter, one can replicate the lines of reasoning in [9] in order to show that $C^*$ is unbounded in $P$ with a double-logarithmic scaling under the condition that there exists an index $i$ such that the tap process $h_{t,i}$ is stationary and ergodic with a finite differential entropy rate.

The rest of the paper is organized as follows. We end the current section by a list of adopted notations. System model, the problem statement and a summary of main results appear in Section II. Section III further explores the lower bound presented in Theorem 1 in Section II. Section IV describes the proposed decoder. Section V is devoted to error analysis where we establish the main result (Theorem 1). Section VI presents achievable rates in the presence of channel state information at the receiver end. The majority of details of the proofs are deferred to the appendices at the end of the paper. Finally, Section VII concludes the paper.

C. Notations

For a real number $x$, $x^+ = \max\{0, x\}$. The Euclidean space of sequences of length $n$ whose entries are real numbers is denoted by $\mathbb{R}^n$. Vectors and sequences are identified by an underline such as $\underline{x}$. Random quantities appear in bold such as $\underline{x}$ and $\underline{\underline{x}}$ with realizations $x$ and $\underline{x}$, respectively. An $m \times n$ matrix whose all entries are zeros is denoted by $0_{m,n}$. We use $\underline{a}_n$ to denote a column vector of length $n$ whose all entries are zeros. The $n \times n$ identity matrix is denoted by $I_n$. The transpose, inverse, trace and determinant of a square matrix $A$ are denoted by $A^T$, $A^{-1}$, tr$(A)$ and det$(A)$, respectively. The entry at the $i^{th}$ row and the $j^{th}$ column of a matrix $A$ is denoted by $A_{i,j}$ or $[A]_{i,j}$. The $i^{th}$ entry of a vector $\underline{x}$ is denoted by $x_i$ or $[\underline{x}]_i$. The 2-norm (Frobenius norm) and the spectral norm (operator norm) of a matrix $A$ are defined by

$$\|A\|_2 := (\text{tr}(A^T A))^{1/2}$$

and

$$\|A\| := \max_{\|\underline{x}\|_2 \leq 1} \|A\underline{x}\|_2 = (\max_{i} \lambda_{\max}(A^T A))^{1/2},$$

respectively, where $\lambda_{\max}(M)$ denotes the largest eigenvalue of a symmetric matrix $M$. The smallest eigenvalue of a symmetric matrix $M$ is denoted by $\lambda_{\min}(M)$. The probability law of a random variable $\underline{x}$ (the induced probability measure on the range of $\underline{x}$) is denoted by $L_{\underline{x}}(\cdot)$ and its mean is denoted by $E[\underline{x}]$. The probability density function (PDF) of a continuous random variable $\underline{x}$ is denoted by $p_{\underline{x}}(\cdot)$ and its differential entropy is denoted by $h(\underline{x})$. The PDF of a Gaussian random vector of length $n$ with zero mean and covariance matrix $\Sigma$ is denoted by

$$p_G(\underline{x}; \Sigma) := \frac{1}{(2\pi)^{n/2} \sqrt{\text{det}(\Sigma)}} \exp\left(-\frac{1}{2} \underline{x}^T \Sigma^{-1} \underline{x}\right).$$

We write $\underline{x} \sim N(\underline{0}_n, \Sigma)$ if $p_G(\underline{x} ; \Sigma) = p_G(\underline{x}; \Sigma)$. The differential entropy of a random vector $\underline{x} \sim N(\underline{0}_n, \Sigma)$ is denoted by $h_G(\Sigma)$ given by

$$h_G(\Sigma) := \frac{1}{2} \log \left( (2\pi e)^n \det(\Sigma) \right),$$

where $\log(\cdot)$ is the logarithm function with base 2. We also recall the definition for a typical set [17]. For $\eta > 0$, positive integer $n$ and an $n \times n$ positive-definite matrix $\Sigma$, the Gaussian typical set $T^{(n)}(\eta) (\Sigma)$ is defined as the set of all $\underline{a} \in \mathbb{R}^n$ such that

$$\left| \frac{1}{n} \log p_G(\underline{a}; \Sigma) + \frac{1}{n} h_G(\Sigma) \right| < \frac{\eta}{2 \log 2}.$$ This simplifies to

$$T^{(n)}(\eta) (\Sigma) = \left\{ \underline{a} \in \mathbb{R}^n : \left| -\frac{1}{n} \Sigma^{-1} \underline{a}^T \underline{a} - 1 \right| < \eta \right\}.$$ (5)

II. System Model and Summary of Results

The stochastic and time-varying Gaussian ISI channel with memory length $k \geq 1$ is defined by

$$y_t = \sum_{i=0}^{k} h_{t,i} x_{t-i} + z_t, \quad t = 1, 2, \cdots.$$ (6)
where \( h_{t,0}, h_{t,1}, \cdots, h_{t,k} \) are the channel coefficients (filter taps) during time slot \( t = 1, 2, \cdots \) and \( x_t, y_t \) and \( z_t \) are the channel input, the channel output and the additive noise at time slot \( t \), respectively. The process \( z_1, z_2, \cdots \) is white Gaussian with zero mean and unit variance. The input process \( x_1, x_2, \cdots \) is subject to the average power constraint

\[
\sum_{t=1}^{n} \mathbb{E}[|x_t|^2] \leq nP,
\]

where \( n \) is the length of the communication period of interest.

The following assumptions are made on the array of channel coefficients:

(i) There is a sequence of numbers \( c = (c_0, c_1, \cdots, c_k) \) and a sequence of nonnegative numbers \( r = (r_0, r_1, \cdots, r_k) \) such that the probability law for \( h_{t,i} \) is supported over the interval \( [c_i - r_i, c_i + r_i] \) for every \( t, i \). We denote the sum of the radii \( r_i \) by \( r_s \), i.e.,

\[
r_s := \sum_{i=0}^{k} r_i.
\]

(ii) The transmitter and the receiver only know the centres \( c_i \) and the radii \( r_i \). The joint distribution of the random variables \( h_{t,i} \) as well as their realizations are unknown to both ends of communication.

(iii) The processes \( h_{t,i} \) and \( z_t \) are independent.

To describe the ISI channel in matrix form, assume a codeword is transmitted during time slots \( t = 1, \cdots, n \). Define

\[
m := n + k
\]

and

\[
\begin{align*}
\mathbf{z} & := [x_1 \ x_2 \ \cdots \ x_n]^T \\
\mathbf{y} & := [y_1 \ y_2 \ \cdots \ y_m]^T \\
\mathbf{z} & := [z_1 \ z_2 \ \cdots \ z_m]^T.
\end{align*}
\]

Then

\[
\mathbf{y} = \mathbf{H}\mathbf{z} + \mathbf{z},
\]

where the \( m \times n \) random channel matrix \( \mathbf{H} \) is given by

\[
H_{i,j} := \begin{cases} 
1 & 0 \leq i-j \leq k \\
0 & \text{otherwise}
\end{cases}
\]

We also define the \( m \times n \) deterministic matrix \( \mathbf{H}_c \) by

\[
[H_c]_{i,j} := \begin{cases} 
c_{i-j} & 0 \leq i-j \leq k \\
0 & \text{otherwise}
\end{cases}
\]

For example, when \( n = 4 \) and \( k = 2 \), we have \( m = 4 + 2 = 6 \) and the matrices \( \mathbf{H} \) and \( \mathbf{H}_c \) are given by

\[
\mathbf{H} = \begin{bmatrix}
h_{1,0} & 0 & 0 & 0 \\
h_{2,0} & 0 & 0 & 0 \\
h_{3,2} & h_{3,1} & h_{3,0} & 0 \\
0 & h_{4,2} & h_{4,1} & h_{4,0} \\
0 & 0 & h_{5,2} & h_{5,1} \\
0 & 0 & 0 & h_{6,2}
\end{bmatrix}
\]

\[
\mathbf{H}_c = \begin{bmatrix}
c_0 & 0 & 0 & 0 \\
c_1 & c_0 & 0 & 0 \\
c_2 & c_1 & c_0 & 0 \\
0 & c_2 & c_1 & c_0 \\
0 & 0 & c_2 & c_1 \\
0 & 0 & 0 & c_2
\end{bmatrix}.
\]

Throughout the paper, we will denote the range of the random matrix \( \mathbf{H} \) by \( \mathcal{H} \), i.e., \( \mathcal{H} \) is the set of all \( m \times n \) matrices \( \mathbf{H} \) such that \( H_{t,j} = 0 \) for \( i-j < 0 \) or \( i-j > k \) and \( |H_{t,j} - c_{i-j}| \leq r_{i-j} \) for \( 0 \leq i-j \leq k \).

The message \( \mathbf{W} \) is uniformly distributed over the set of indices \( \{1, 2, \cdots, 2^{nR}\} \) where \( R \) is the transmission rate. The encoder maps \( \mathbf{W} \) to a codeword \( \mathbf{z} \) of length \( n \). This codeword is then transmitted over the channel in (11) during time slots \( t = 1, \cdots, n \). The decoder receives the vector \( \mathbf{y} \) and generates the estimate \( \hat{\mathbf{W}} \) for \( \mathbf{W} \). The probability of decoding error is denoted by

\[
p_{e,n} := \Pr(\hat{\mathbf{W}} \neq \mathbf{W}).
\]

We say a transmission rate \( R \) is achievable if there exists a sequence of encoder-decoder pairs such that

\[
\lim_{n \to \infty} p_{e,n} = 0.
\]

The supremum of all achievable rates is denoted by \( C^* = C^*(\mathcal{C}, \mathcal{P}) \) and referred to as the capacity of the stochastic and time-varying ISI channel with parameters \( \mathcal{C} \) and \( \mathcal{P} \) under a maximum average power of \( P \).

Let

\[
f(\omega) := \sum_{l=0}^{k} c_l e^{-\pi T l^2} \delta(\omega),
\]

be the discrete Fourier transform of the sequence \( \mathcal{C} \). We refer to \( f(\omega) \) as the central channel frequency response. In the special case where the ISI channel is deterministic and time-invariant, i.e., \( r_i = 0 \) for every \( i \), it is well-known\(^1\) that the capacity, denoted in this case by \( C_{0}^* := C^*(\mathcal{C}_0, \mathcal{P}) \), is given by

\[
C_{0}^* = \frac{1}{4\pi} \int_{0}^{2\pi} \log[\max(\Theta, |f(\omega)|^2)] d\omega,
\]

where the parameter \( \Theta \) solves

\[
\frac{1}{2\pi} \int_{0}^{2\pi} (\Theta - |f(\omega)|^{-2})^+ d\omega = P.
\]

The main contributions of this paper are a lower bound (achievability result) on \( C^* \) for arbitrary \( \mathcal{C}, \mathcal{P} \) and an upper bound (converse result) on \( C^* \). To present these results, fix a sequence of positive-definite matrices \( \Sigma_n \) such that\(^2\)

\[
\text{tr}(\Sigma_n) \leq nP
\]

and

\[
\lim_{n \to \infty} \frac{1}{n} \lambda_{\max}(\Sigma_n) = 0.
\]

\(^1\) See [2] and the references therein.

\(^2\) As a non-example, the matrices \( \Sigma_n \) whose all diagonal entries are \( P \) and whose all entries off the main diagonal are \( rP \) for some \( r \neq 0 \) satisfy (20), but they do not satisfy (21).
Define
\[ \alpha := \min_{\omega \in [0,2\pi]} |f(\omega)|, \quad \beta := \max_{\omega \in [0,2\pi]} |f(\omega)|, \] (22)
and
\[ \gamma_n := r_s (r_s + 2\beta) \lambda_{\text{max}}(\Sigma_n) \left( \frac{1}{1 + \alpha^2 \lambda_{\text{min}}(\Sigma_n)} \right)^2. \] (23)

We are ready to present the achievability result. 

**Theorem 1**: Let \( \Sigma_n \) be a sequence of positive-definite matrices that satisfies both (20) and (21). Assume \( r_s \) is small enough such that \( \gamma_n < 1 \) for all \( n \). Then \( C^* \geq C^*_{LB} \) where
\[
C^*_{LB} := \liminf_{n \to \infty} \left( \frac{1}{2n} \log (I_n + H_c^T H_c \Sigma_n) - \log \left( 1 + (k + 1) \left\| \frac{\text{tr}(\Sigma_n)}{m + n} \right\| \right) - \delta_n \right) \] (24)
and
\[ \delta_n := -\frac{1}{2} \log (1 - \gamma_n). \] (25)

**Proof**: The proof is detailed in Section V. \( \square \)

The lower bound \( C^*_{LB} \) is achieved by a mismatched Gaussian joint-typicity decoder that is tuned to the matrix \( H_c \) rather than the unknown channel matrix \( H \). In the absence of channel variations, i.e., when \( r_0, r_1, \ldots, r_k \) are all equal to zero, one only has the first term \( \frac{1}{2n} \log \det (I_n + H_c^T H_c \Sigma_n) \) in (24). The second term \( \log \left( 1 + (k + 1) \left\| \frac{\text{tr}(\Sigma_n)}{m + n} \right\| \right) \) and the third term \( \delta_n \) in (24) come to life in the presence of channel variations. They appear in the study of the so-called type I and type II error probabilities, respectively. The details are given in Section V.

It is shown in the next section that the lower bound \( C^*_{LB} \) in Theorem 1 saturates at a positive level as \( P \) increases regardless of \( k \geq 0 \). We emphasize that \( C^* \) is achievable under any joint distribution (not just the I.I.D. model or the stationary and ergodic model) for the random array \( h_{t,i} \in [c_i - r_i, c_i + r_i] \). The question rises if the capacity \( C^* \) is unbounded in terms of \( P \) at least under specific conditions on the distribution of the array of channel taps. By eliminating ISI through zero padding\(^3\) and following an idea put forth in [8] in the context of memoryless Rayleigh fading channels, it is shown in Appendix A that \( C^* \) is unbounded in terms of \( P \) if there exists an index \( i = 0, 1, \ldots, k \) such that \( r_i < \frac{1}{2} |c_i| \). Moreover, one can replicate the lines of reasoning in [9] to show that \( C^* \) scales double-logarithmically in terms of \( P \) if there is an index \( i = 0, 1, \ldots, k \) such that the channel tap process \( h_{t,i} \) for \( t = 1, 2, \ldots \) is stationary and ergodic with a finite differential entropy rate. Once again, intermittent zero padding is required in order to draw this conclusion.

Next, we present a "partial converse" result which provides an upper bound on \( C^* \) under additional conditions on the array of channel taps. For every \( 0 \leq i \leq k \), let \( b_i \) be the \( n \times 1 \) vector whose entries are the \( i \)th channel taps that appear in the formulation of the random matrix \( H \) in (12). These entries are the \( n \) taps \( h_{t,i} \) at times \( t = i + 1, i + 2, \ldots, i + n \), i.e.,
\[ b_i := [h_{i+1,i} \ h_{i+2,i} \ \cdots \ h_{i+n,i}]^T. \] (26)

In fact, the matrix \( H \) has \( k + 1 \) nonzero diagonals of random entries. The vectors \( b_0, b_1, \ldots, b_k \) carry the entries on those \( k + 1 \) diagonals. For example, let \( n = 4 \) and \( k = 2 \). The matrix \( H \) is shown in (14). The three vectors \( b_0, b_1 \) and \( b_2 \) are given by
\[ b_0 = [h_{1,0} \ h_{2,0} \ h_{3,0} \ h_{4,0}]^T, \] (27)
\[ b_1 = [h_{2,1} \ h_{3,1} \ h_{4,1} \ h_{5,1}]^T, \] (28)
and
\[ b_2 = [h_{3,2} \ h_{4,2} \ h_{5,2} \ h_{6,2}]^T. \] (29)

We impose the next two additional conditions on the array of channel taps:
1) The vectors \( b_0, b_1, \ldots, b_k \) are independent. This means that different taps vary independently of each other.
2) Each of the vectors \( b_0, b_1, \ldots, b_k \) has a density with a finite differential entropy. We define
\[ \rho_n := \max_{i=0,1,\ldots,k} \frac{\gamma(b_i)_n}{n}. \] (30)

**Theorem 2**: Assume the array of channel taps satisfies conditions (1) and (2) given in above. Let \( C_n = \{x_{n,1}, \ldots, x_{n,2n-1}\} \) for \( n = 1, 2, \ldots \) be a sequence of codebooks with rate \( R \) and vanishingly small probability of error that satisfy the average power constraint in (7), i.e.,
\[ \frac{1}{2nR} \sum_{i=1}^{2nR} \|x_{n,i}\|^2 \leq nP, \quad n = 1, 2, \ldots. \] (31)

Moreover, let every symbol in every codeword \( x_{n,i} \) be nonzero. Then
\[ R \leq C^*_{UB} := \liminf_{n \to \infty} \frac{1}{2nR} \sum_{i=1}^{2nR} F_n(x_{n,i}), \] (32)
where the function \( F_n : \mathbb{R}^n \to (0, \infty) \) is defined by
\[ F_n(x) := \frac{1}{2} \log \left( 2(\beta + r_s)^2 \left\| x \right\|^2 + 2 \right) - \frac{1}{2} \log \left( 1 + \frac{2^2 r_s}{2\pi e} \prod_{j=1}^{n} s_{j}^{2/n} \right). \] (33)

Here, \( \rho_n \) is defined in (30) and \( x_j \) is the \( j \)th symbol of the vector \( x \in \mathbb{R}^n \).

**Proof**: See Appendix B. \( \square \)

The first term on the right side of (33) depends on the arithmetic mean of the squared symbols of \( x \) and the second term depends on the geometric mean of those squared symbols. Let us define
\[ \tau := \liminf_{n \to \infty} \frac{1}{2nR} \sum_{i=1}^{2nR} \frac{1}{2} \log \left( 1 + \frac{2^2 r_s}{2\pi e} \prod_{j=1}^{n} s_{j}^{2/n} \right). \] (34)

The upper bound \( C^*_{UB} \) can be further relaxed as shown in (35), shown at the bottom of the next page, where \( (a) \) is due to the inequality \( \lim inf_{n \to \infty} (a_n - b_n) \leq \lim inf_{n \to \infty} a_n - \lim inf_{n \to \infty} b_n \) for every two sequences \( a_n, b_n \), \( (b) \) is due to

\(^{3}\)ISI is avoided by sending a sequence of \( k \) zeros after sending each code symbol.
Jensen’s inequality and concavity of $\log(\cdot)$ and (c) is due to the power constraint in (31). If a sequence of codebooks is such that $\tau = \frac{1}{2} \log P + O(1)$ where $O(1)$ does not scale with $P$, then a direct consequence of Theorem 2 is that every rate achieved by that sequence of codebooks is bounded from above by an expression that does not scale with $P$. This in particular true if (i) $\liminf_{n \to \infty} n \rho_n > -\infty$ and (ii) the absolute value of every symbol in every codeword is at least $a \sqrt{P}$ for some constant $a > 0$. The former certainly holds if the process $h_{t,i}$ for $t = 1, 2, \cdots$ is stationary with a finite differential entropy rate for at least one choice of $i = 0, 1, \cdots, k$.

### III. Exploring The Lower Bound in Theorem 1

The matrix $H_c^T H_c$ is a symmetric banded Toeplitz matrix whose $(i,j)$-entry is given by

$$[H_c^T H_c]_{i,j} = \begin{cases} \sum_{k=0}^{\min(i,j)} c_{i-k, j-k} & \text{for } |i-j| \leq k \\ 0 & \text{otherwise.} \end{cases}$$

Let $H_c^T H_c = U \Lambda U^T$ be the eigenvalue decomposition of $H_c^T H_c$ where $U$ is an $n \times n$ orthogonal matrix and the diagonal matrix $\Lambda$ carries the eigenvalues

$$\lambda_{1,n} \leq \cdots \leq \lambda_{n,n}$$

of $H_c^T H_c$ on its diagonal. We choose

$$\Sigma_n = U^T D U,$$

where

$$D = \text{diag}(d_{n,1}, \cdots, d_{n,n}),$$

and $d_{n,1}, \cdots, d_{n,n}$ satisfy

$$\sum_{i=1}^{n} d_{n,i} \leq n P$$

and

$$\lim_{n \to \infty} \frac{1}{n} \max_{1 \leq i \leq n} d_{n,i} = 0.$$  

The conditions in (40) and (41) are the ones in (20) and (21), respectively. Then $C_{LB}^*$ can be written as

$$C_{LB}^* = \liminf_{n \to \infty} \left( \frac{1}{2n} \sum_{i=1}^{n} \log(1 + \lambda_{n,i} d_{n,i}) \right. - \log \left( 1 + (k + 1) \|P\|_2^2 \frac{\sum_{i=1}^{n} d_{n,i}}{n + n} \right) / 2 \left) - \frac{1}{2} \right).$$

Since $\lambda_{n,i}$ is the coefficient of $d_{n,i}$ in $\sum_{i=1}^{n} \log(1 + \lambda_{n,i} d_{n,i})$, we let

$$0 < d_{n,1} \leq d_{n,2} \leq \cdots \leq d_{n,n},$$

to ensure $C_{LB}^*$ is the largest. As mentioned earlier, we require $\Sigma_n$ be positive-definite and hence, all eigenvalues of $\Sigma_n$ must be positive. The term $\delta_n$ in (25) is given by

$$\delta_n = -\frac{1}{2} \log \left( 1 + \frac{r_s (r_s + 2 \beta d_{n,n})}{1 + \alpha^2} \right).$$

The lower bound in (42) can be maximized over all choices of $d_{n,1}, \cdots, d_{n,n}$ that satisfy the conditions in (40), (41) and (43). For simplicity, we will look into the problem of maximizing

$$\frac{1}{2n} \sum_{i=1}^{n} \log(1 + \lambda_{n,i} d_{n,i}) - \log(1 + (k + 1) \|P\|_2^2 (\sum_{i=1}^{n} d_{n,i}) - \frac{1}{2} \right).$$

The former certainly does not scale with $n$, while the latter scales like $\frac{1}{2} \log \left( 1 + \frac{r_s (r_s + 2 \beta d_{n,n})}{1 + \alpha^2} \right)$. The proof of this proposition is established.

### Proposition 1

Let $\Theta_1$ be the solution to

$$\frac{1}{2\pi} \int_0^{2\pi} (\Theta - |f(\omega)|^{-2})^t d\omega = P,$$

and $\Theta_2$ be the solution to

$$\frac{1}{2\pi} \int_0^{2\pi} (\Theta - |f(\omega)|^{-2})^t d\omega = 2 \Theta - \frac{2}{k + 1} \|P\|_2^2.$$

For $i = 1, 2$ define

$$I_i := \frac{1}{2\pi} \int_0^{2\pi} (\Theta - |f(\omega)|^{-2})^t d\omega$$

and

$$\phi_i := -\frac{1}{2} \log \left( 1 + \frac{r_s (r_s + 2 \beta d_{\max,i})}{1 + \alpha^2} \right).$$

and

$$C_{LB,i}^* := \frac{1}{4\pi} \int_0^{2\pi} \log(\max(\Theta, |f(\omega)|^2, 1)) d\omega$$

and

$$\phi_i := -\frac{1}{2} \log \left( 1 + \frac{k + 1}{2} \|P\|_2^2 I_i - \phi_i \right).$$

Then $C^*$ is bounded from below by $C_{LB,*}$. If $I_1 \geq I_2$, then $C_{LB,2}^*$ is also a lower bound on $C^*$.

### Proof

See Appendix C.

We see that if $r_0 = r_1 = \cdots = r_k = 0$, the lower bound $C_{LB,1}$ reduces to the capacity $C_0^*$ in (18). An immediate corollary of this proposition is an upper bound on the reduction in the channel capacity compared to when all radii $r_i$ are zeros.

$$C_{UB}^* \leq \liminf_{n \to \infty} \left( \frac{1}{2n} \sum_{i=1}^{2n} \log \left( 2(\beta + r_s)^2 \|x_{n,i}\|_2^2 + 2 \right) - \tau \right.$$

$$\leq \liminf_{n \to \infty} \left( \frac{1}{2} \log \left( 2(\beta + r_s)^2 \|x_{n,i}\|_2^2 + 2 \right) - \tau \right.$$

$$\leq \frac{1}{2} \log \left( 2(\beta + r_s)^2 P + 2 \right) - \tau.$$

(35)
Corollary 1: We have

$$C_0^* - C^* \leq \log \left( 1 + \frac{k+1}{2} \parallel r \parallel_2^2 P \right) + \phi_1,$$  \hspace{1cm} (51)

where $\phi_1$ is given in (49).

Proof: Since $\Theta_1$ solves the equation in (19), then $C_0^* = \frac{1}{2\pi} \int_0^{2\pi} \max(\Theta_1 |f(\omega)|^2, 1) d\omega$. Hence, the lower bound $C_{LB,1}^*$ in (50) can be written as $C_{LB,1}^* = C_0^* - \log(1 + \frac{k+1}{2} \parallel r \parallel_2^2 P) - \phi_1$ verifying (51).

We note that $\phi_1$ depends on the radii $r_i$ only through their sum $r_s$. Since $\parallel r \parallel_2^2 \leq \sum_{i=0}^k r_i^2 = r_s^2$, one can loosen the upper bound on $C_0^* - C^*$ in (51) in order to write it entirely in terms of $r_s$ as

$$C_0^* - C^* \leq \log \left( 1 + \frac{k+1}{2} r_s^2 P \right) - \frac{1}{2} \log (1 - \frac{r_s (r_s + 2\beta d_{\min,1})}{1 + \alpha^2 d_{\min,1}}).$$  \hspace{1cm} (52)

Figure 1 in panel (a) presents plots of the upper bound on $C_0^* - C^*$ given in (52) in terms of $r_s$ for several values of $P$ in an ISI channel with $k = 2$, $c_0 = 1$ and $c_1 = 0.5$. Panel (b) presents similar plots in an ISI channel with $k = 4$, $c_0 = 1$, $c_1 = 0.5$, $c_2 = 0.4$, $c_3 = 0.3$ and $c_4 = 0.2$. In both cases, the incurred loss in capacity is less than 0.5 bits/sec/Hz regardless of $r_s \leq 0.01$.

Next, we look into the solutions $\Theta_1$ and $\Theta_2$ to the equations (45) and (46), respectively. During the rest of this section, we assume $f(\omega)$ is everywhere nonzero and hence, $\alpha > 0$. Let us define

$$J := \frac{1}{2\pi} \int_0^{2\pi} |f(\omega)|^{-2} d\omega.$$  \hspace{1cm} (53)

It is easy to find $\Theta_1$ when $P$ is sufficiently large. To see this, assume $\Theta_1 \geq \max_{0 \leq \omega \leq 2\pi} |f(\omega)|^{-2} = \frac{1}{\beta}$. Then $\frac{1}{2\pi} \int_0^{2\pi} (\Theta_1 - |f(\omega)|^{-2})^+ d\omega = \Theta_1 - J$ and (45) gives $\Theta_1 = P + J$. If this value happens to be larger than or equal to $\frac{1}{\alpha^2}$, it must be the unique solution to (45), i.e.,

$$P \geq \frac{1}{\alpha^2} - J \Rightarrow \Theta_1 = P + J.$$  \hspace{1cm} (54)

Then the lower bound $C_{LB,1}^*$ is written as

$$C_{LB,1}^* = \frac{1}{2\pi} \int_0^{2\pi} |f(\omega)| d\omega + \frac{1}{2} \log(P + J) - \log \left( 1 + \frac{k+1}{2} \parallel r \parallel_2^2 P \right) - \phi_1.$$  \hspace{1cm} (55)

It is also easy to find $\Theta_2$ when $\parallel r \parallel_2^2$ is sufficiently small. A similar argument that led to (54) gives

$$\frac{2}{k+1} \parallel r \parallel_2^2 \geq \frac{1}{\alpha^2} + J \Rightarrow \Theta_2 = \frac{2}{k+1} \frac{1}{\parallel r \parallel_2^2} - J.$$  \hspace{1cm} (56)

Then the lower bound $C_{LB,2}^*$ is written as

$$C_{LB,2}^* = \frac{1}{2\pi} \int_0^{2\pi} |f(\omega)| d\omega - \log((k+1) \parallel r \parallel_2^2) - \frac{1}{2} \log \left( \frac{2}{k+1} \frac{1}{\parallel r \parallel_2^2} - J \right) - \phi_2.$$  \hspace{1cm} (57)

The lower bound $C_{LB,2}^*$ does not depend on $P$. The lower bound $C_{LB,1}^*$ reaches a maximum and eventually decreases as $P$ increases. In fact, if one maximizes the expression $\frac{1}{2} \log(P + J) - \log(1 + \frac{k+1}{2} \parallel r \parallel_2^2 P)$ as a function of $P$ in the given expression for $C_{LB,1}^*$ in (55), the maximum occurs at $P = \frac{2}{k+1} \frac{1}{\parallel r \parallel_2^2} - 2J$ and the resulting maximum value is precisely $-\log((k+1) \parallel r \parallel_2^2) - \frac{1}{2} \log \left( \frac{2}{k+1} \frac{1}{\parallel r \parallel_2^2} - J \right)$ which appears in the given expression for $C_{LB,2}^*$ in (57). Moreover, recall that $C_{LB,2}^*$ is a lower bound on $C^*$ only when $I_1 \geq I_2$.

This inequality gives $P \geq 2\left( \frac{2}{k+1} \frac{1}{\parallel r \parallel_2^2} - J \right) - \frac{2}{k+1} \frac{1}{\parallel r \parallel_2^2} = \frac{2}{k+1} \frac{1}{\parallel r \parallel_2^2} - 2J$. Therefore, $\frac{2}{k+1} \frac{1}{\parallel r \parallel_2^2} - 2J$ is also the smallest value for $P$ beyond which $C_{LB,2}^*$ is a valid lower bound on $C^*$. We refer to this value of $P$ as the saturation power and denote it by $P_{sat}$, i.e.,

$$P_{sat} := \frac{2}{k+1} \frac{1}{\parallel r \parallel_2^2} - 2J.$$  \hspace{1cm} (58)

Figure 2 in Panel (a) presents plots of the capacity $C_0^*$ and the lower bounds $C_{LB,1}^*$ and $C_{LB,2}^*$ in terms of $P$ in an ISI channel with parameters $k = 2$, $c_0 = 1$, $c_1 = c_2 = 0.5$ and $r_0 = r_1 = r_2 = 0.001$. The saturation power is $P_{sat} \approx 53.47$ dB. Figure 2 in Panel (b) presents similar
plots in an ISI channel with parameters $k = 2$, $c_0 = 1$, $c_1 = c_2 = 0.5$ and $r_0 = r_1 = r_2 = 0.01$. The saturation power is $P_{sat} \approx 53.47$ dB. Panel (b) belongs to an ISI channel with $k = 2$, $c_0 = 1$, $c_1 = c_2 = 0.5$ and $r_0 = r_1 = r_2 = 0.01$. The saturation power is $P_{sat} \approx 33.46$ dB.

We end this section with the next result which gives an upper bound on $C_0^* - C^*$ before saturation occurs.

**Corollary 2:** Assume $\frac{2}{k+1} \frac{1}{\left\| x \right\|^2} \geq \frac{1}{\alpha^2} + J$. Then regardless of the value of $P \leq P_{sat}$, the capacity loss is bounded as

$$C_0^* - C^* \leq 1 - \frac{1}{2} \log \left( 1 - \frac{r_s(r_s + 2\beta)}{\alpha^2} \right).$$  \hspace{1cm} (59)

**Proof:** By (51) in Corollary 1,

$$C_0^* - C^* \leq \log \left( 1 + \frac{k+1}{2} \left\| x \right\|^2 P \right)
- \frac{1}{2} \log \left( 1 - \frac{r_s(r_s + 2\beta)}{\alpha^2} d_{max,1} \right).$$  \hspace{1cm} (60)

Since $P \leq P_{sat} = \frac{k+1}{2} \frac{1}{\left\| x \right\|^2} - 2J$, then

$$\log \left( 1 + \frac{k+1}{2} \left\| x \right\|^2 P \right) \leq \log \left( 1 + \frac{k+1}{2} \left\| x \right\|^2 P_{sat} \right)
= \log (2 - (k+1) \left\| x \right\|^2 J)
\leq 1.$$  \hspace{1cm} (61)

Therefore,

$$C_0^* - C^* \leq 1 - \frac{1}{2} \log \left( 1 - \frac{r_s(r_s + 2\beta)}{\alpha^2} d_{max,1} \right).$$  \hspace{1cm} (62)

The condition $\frac{2}{k+1} \frac{1}{\left\| x \right\|^2} \geq \frac{1}{\alpha^2} + J$ gives $P_{sat} \geq \frac{1}{\alpha^2} - J$.

We study the cases $\frac{1}{\alpha^2} - J \leq P \leq P_{sat}$ and $P \leq \frac{1}{\alpha^2} - J$ separately.

1) First, let $P \geq \frac{1}{\alpha^2} - J$. Then $\Theta_1 = P + J$ by (54) and we have $d_{min,1} = (P + J - \alpha^{-2})^+ = P + J - \alpha^{-2}$ and $d_{max,1} = (P + J - \beta^{-2})^+ = P + J - \beta^{-2}$. Hence,

$$d_{max,1} = \frac{P + J - \beta^{-2}}{1 + \alpha^2 (P + J - \alpha^{-2})} = \frac{P + J - \beta^{-2}}{\alpha^2 (P + J)}.$$  \hspace{1cm} (63)

Then (62) and (63) give the desired result.

2) Next, let $P \leq \frac{1}{\alpha^2} - J$. We note that the solution $\Theta_1$ to (45) is nondecreasing in terms of $P$. Since the condition $P \geq \frac{1}{\alpha^2} - J$ gives $\Theta_1 = P + J$, then $P \leq \frac{1}{\alpha^2} - J$ implies $\Theta_1 \leq P + J$. Hence, $d_{min,1} = (\Theta_1 - \alpha^{-2})^+ \leq (P + J - \alpha^{-2})^+ = 0$ and $d_{max,1} = (\Theta_1 - \beta^{-2})^+ \leq P + J - \beta^{-2} \leq \alpha^{-2} - \beta^{-2}$. Once again the inequality in (63) holds and the desired result follows.

As an example, consider the ISI channel with $k = 2$, $c_0 = 1$, $c_1 = c_2 = 0.5$ and $r_0 = r_1 = r_2 = 0.01$. Then $r_s = 0.03$, $\alpha = 0.4677$, $\beta = 2$ and Corollary 2 guarantees that $C_0^* - C^* \leq 1.5803$ bits/sec/Hz regardless of $P \leq P_{sat} \approx 33.46$ dB.

### IV. THE DECODER STRUCTURE

The proposed decoder applies Gaussian joint-typicality decoding tuned to the matrix $H_c$ defined in (13). Denote the codewords by the independent random vectors $\bar{x}_i \sim N(0, \Sigma_n)$ for $i = 1, 2, \cdots, 2^nR$ where $\Sigma_n$ is a positive-definite matrix. Recall the definition of Gaussian typical sets in (5). The Gaussian joint-typicality decoder looks for the unique index $i \in \{1, \cdots, 2^nR\}$ such that

$$\bar{x}_i \in T_{\epsilon}^{\eta}(\Sigma_n), \quad \bar{w}_i \in T_{\eta}^{\eta+m+n}(\Xi_n),$$  \hspace{1cm} (64)

where $\epsilon, \eta > 0$ are constants and $\bar{w}_i$ and $\Xi_n$ are defined by

$$\bar{w}_i := \left[ \bar{x}_i^T \quad \bar{y}_i^T \right]^T, \quad \Xi_n := \left[ \Sigma_m \quad \Sigma_nH_c^T \right] = \left[ \begin{array}{cc} \Sigma_m & \Sigma_nH_c^T \\ H_c^T & I_m + H_c\Sigma_nH_c^T \end{array} \right].$$  \hspace{1cm} (65)

The matrix $\Sigma_n$ is the covariance matrix for a vector $[\bar{x}_i^T \bar{y}_i^T]^T$ where $\bar{y}_i = H_c \bar{x}_i + \bar{z}_i$ and $\bar{x}_i$ and $\bar{z}_i$ are independent $N(0, \Sigma_n)$ and $N(0, I_m)$ random vectors, respectively. To construct the set $T_{\eta}^{\eta+m+n}(\Xi_n)$, one needs to compute $\Xi_n^{-1}$. It is easy to see that $\Xi_n^{-1}$ admits a closed form given by

$$\Xi_n^{-1} = \left[ \begin{array}{cc} \Sigma_n^{-1} + H_c^T\Sigma_m & -H_c^T \\ -H_c & I_m \end{array} \right].$$  \hspace{1cm} (66)
It will be useful in the course of our computations to note that\footnote{This follows by the Schur’s complement formula for calculating the determinant of partitioned matrices [12].} \[
\det(\Xi_n) = \det(\Sigma_n), \quad (67)
\]
regardless of $H_c$.

V. ERROR ANALYSIS

Two types of error are distinguished:

(i) The transmitted codeword does not satisfy the decoding rule in (64). We refer to this as the type I error and denote it by $Error^{(I)}$.

(ii) A codeword different from the transmitted codeword satisfies (64). We refer to this as the type II error and denote it by $Error^{(II)}$.

Then
\[
\rho_{e,n} \leq \text{Pr}(Error^{(I)}) + \text{Pr}(Error^{(II)}). \quad (68)
\]

In the following, we examine the two terms on the right side of (68) separately.

A. The Probability of the Type I Error

Without loss of generality, assume $\omega_1$ is the transmitted codeword, i.e., \(y = H\omega_1 + \zeta\). Then
\[
\text{Pr}(Error^{(I)}) \leq \text{Pr}(\omega_1 \notin T_\tau^{(n)}(\Sigma_n)) + \text{Pr}(\omega_1 \notin T_\eta^{(m+n)}(\Xi_n)). \quad (69)
\]

For every $\varepsilon > 0$, the first term on the right side of (69) tends to zero when $n$ grows due to Theorem 5 in [10]. The second term on the right side (69) is studied in the next proposition.

Proposition 2: Assume
\[
\eta > \eta_n := (k + 1)\|x\|^2 \frac{\text{tr}(\Sigma_n)}{m + n}. \quad (70)
\]

Then
\[
\text{Pr}(\omega_1 \notin T_\eta^{(m+n)}(\Xi_n)) \leq C_n((m + n)^2(\eta - \eta_n)^2), \quad (71)
\]

where
\[
C_n := 2m + 2n + 8(k + 1)nP\|x\|^2 + 2nPr^4\lambda_{max}(\Sigma_n). \quad (72)
\]

Proof: The event $\omega_1 \notin T_\eta^{(m+n)}(\Xi_n)$ is equivalent to $\|\frac{1}{m+n}w_1^T\Xi_n^{-1}w_1 - 1\| \geq \eta$. Let us study the term $w_1^T\Xi_n^{-1}w_1$. We write
\[
w_1 = A\nu, \quad (73)
\]

where $A$ and $\nu \sim N(0_{m+n}, I_{m+n})$ are given by\footnote{Here, $\Sigma_n^{1/2}$ is a positive-definite matrix whose square is $\Sigma_n$.}
\[
A := \begin{bmatrix} \Sigma_n^{1/2} & 0_{m,n} \\ \Sigma_n^{1/2} & I_m \end{bmatrix}, \quad \nu := \begin{bmatrix} \Sigma_n^{-1/2} \omega_1^T \\ \omega_1 \end{bmatrix}^T. \quad (74)
\]

Then
\[
w_1^T\Xi_n^{-1}w_1 = \nu^T\Phi\nu, \quad (75)
\]

where $\Phi$ is given by
\[
\Phi := A^T\Xi_n^{-1}A. \quad (76)
\]

Substituting the expression given for $\Xi_n^{-1}$ in (66) and performing simple algebra, we find that
\[
\Phi = \begin{bmatrix} I_n + \Sigma_n^{1/2}E^T\Sigma_n^{1/2} & \Sigma_n^{1/2}E^T \\ \Sigma_n^{1/2}E & I_m \end{bmatrix}, \quad (77)
\]

where the “error matrix” $E$ is defined by
\[
E := H - H_c, \quad (78)
\]
i.e., $E$ is the difference between the actual channel matrix $H$ and the matrix $H_c$. By (75)
\[
\text{Pr}(\omega_1 \notin T_\eta^{(m+n)}(\Xi_n)) = \text{Pr}\left(\frac{1}{m+n}\nu^T\Phi\nu - 1 \geq \eta\right). \quad (79)
\]

We continue to further bound the term on the right side of (79). Recall the range for the random matrix $H$ was denoted by $\mathcal{H}$ in Section II. For $H \in \mathcal{H}$, we denote the corresponding realizations for $E$ and $\Phi$ by $E$ and $\Phi$, respectively. Let us write the probability $\text{Pr}\left((\nu^T\Phi\nu - (m + n)) \geq (m + n)\eta\right)$ as in (80), shown at the bottom of the next page, where $\mathcal{L}_H(\cdot)$ is the probability law for $H$. We bound the term $\text{Pr}\left((\nu^T\Phi\nu - (m + n)) \geq (m + n)\eta\right)$ from above as given in (81), shown at the bottom of the next page, where $a$ is due to independence of $H$ and $\nu$, (b) follows by adding and subtracting $E[\nu^T\Phi\nu]$ and applying the triangle inequality and (c) is due to
\[
E[\nu^T\Phi\nu] = \text{tr}(\Phi) = m + n + \text{tr}(\Sigma_n^{1/2}E^T\Sigma_n^{1/2}). \quad (82)
\]

But,
\[
\text{tr}(\Sigma_n^{1/2}E^T\Sigma_n^{1/2}) = \text{tr}(E^T E\Sigma_n) \quad (a) \geq \text{tr}(E^T E\Sigma_n) \quad (b) \geq |\text{tr}(E^T E\Sigma_n)| \quad (c) \geq \sum_{i=1}^n |E^T E\Sigma_n|_{i,i} \quad (d) \leq \sum_{i=1}^n \sum_{j=1}^n |E^T E|_{i,j} |\Sigma_n|_{j,i} \quad (e) \leq \sum_{i=1}^n \sum_{j=1}^n (|E^T E|_{i,j} |\Sigma_n|_{i,j} |\Sigma_n|_{j,i})^{1/2}, \quad (83)
\]

where (a) is due to the identity $\text{tr}(M_1 M_2) = \text{tr}(M_2 M_1)$, (b) is due to $\text{tr}(E^T E\Sigma_n) \geq 0$, (c) is due to the triangle inequality and (d) is due to the fact that $\Sigma_n$ is a positive-definite matrix and hence, each of its central $2 \times 2$ minors are nonnegative, i.e., $|\Sigma_n|_{i,i} |\Sigma_n|_{j,j} - |\Sigma_n|_{i,j}^2 \geq 0$. Moreover,
\[
|E^T E|_{i,j} \leq \sum_{i=1}^m |E_{i,i} E_{i,j}| \quad (a) \leq \sum_{i=1}^m |E_{i,i}| |E_{i,j}| \quad (b).
\]
The right side of (88) does not depend on the choice of \( H \in \mathcal{H} \). By (79), (80) and (88), we arrive at the promised bound in (71).

By the previous proposition, we see that if both conditions
\[
\lim_{n \to \infty} \frac{1}{\eta} \lambda_{\max}(\Sigma_n) = 0 \quad \text{and} \quad \eta > \limsup_{n \to \infty} \eta_n
\]
hold, then
\[
\lim_{n \to \infty} \Pr\left( \mathbf{w}_1 \notin T^{(m+n)}(\Sigma_n) \right) = 0. \tag{89}
\]

\( \eta \) is due to \( \eta > \eta_n \) and Chebyshev’s inequality and (b) is due to \( \text{Var}(\mathbf{t}_m) = 2t(\Phi^2) \). This formula relies on the fact that the right side of (88) does not depend on the choice of \( H \in \mathcal{H} \). By (79), (80) and (88), we arrive at the promised bound in (71).

By the previous proposition, we see that if both conditions
\[
\lim_{n \to \infty} \frac{1}{\eta} \lambda_{\max}(\Sigma_n) = 0 \quad \text{and} \quad \eta > \limsup_{n \to \infty} \eta_n
\]
hold, then
\[
\lim_{n \to \infty} \Pr\left( \mathbf{w}_1 \notin T^{(m+n)}(\Sigma_n) \right) = 0. \tag{89}
\]

\( \eta \) is due to \( \eta > \eta_n \) and Chebyshev’s inequality and (b) is due to \( \text{Var}(\mathbf{t}_m) = 2t(\Phi^2) \). This formula relies on the fact that the right side of (88) does not depend on the choice of \( H \in \mathcal{H} \). By (79), (80) and (88), we arrive at the promised bound in (71).

By the previous proposition, we see that if both conditions
\[
\lim_{n \to \infty} \frac{1}{\eta} \lambda_{\max}(\Sigma_n) = 0 \quad \text{and} \quad \eta > \limsup_{n \to \infty} \eta_n
\]
hold, then
\[
\lim_{n \to \infty} \Pr\left( \mathbf{w}_1 \notin T^{(m+n)}(\Sigma_n) \right) = 0. \tag{89}
\]

\( \eta \) is due to \( \eta > \eta_n \) and Chebyshev’s inequality and (b) is due to \( \text{Var}(\mathbf{t}_m) = 2t(\Phi^2) \). This formula relies on the fact that the right side of (88) does not depend on the choice of \( H \in \mathcal{H} \). By (79), (80) and (88), we arrive at the promised bound in (71).

By the previous proposition, we see that if both conditions
\[
\lim_{n \to \infty} \frac{1}{\eta} \lambda_{\max}(\Sigma_n) = 0 \quad \text{and} \quad \eta > \limsup_{n \to \infty} \eta_n
\]
hold, then
\[
\lim_{n \to \infty} \Pr\left( \mathbf{w}_1 \notin T^{(m+n)}(\Sigma_n) \right) = 0. \tag{89}
\]

\( \eta \) is due to \( \eta > \eta_n \) and Chebyshev’s inequality and (b) is due to \( \text{Var}(\mathbf{t}_m) = 2t(\Phi^2) \). This formula relies on the fact that the right side of (88) does not depend on the choice of \( H \in \mathcal{H} \). By (79), (80) and (88), we arrive at the promised bound in (71).

By the previous proposition, we see that if both conditions
\[
\lim_{n \to \infty} \frac{1}{\eta} \lambda_{\max}(\Sigma_n) = 0 \quad \text{and} \quad \eta > \limsup_{n \to \infty} \eta_n
\]
hold, then
\[
\lim_{n \to \infty} \Pr\left( \mathbf{w}_1 \notin T^{(m+n)}(\Sigma_n) \right) = 0. \tag{89}
\]
where \( \mathcal{L}_H(\cdot) \) is the probability law for \( H \) and \( \mathcal{H} \) is the range for the random matrix \( H \) as defined in Section II. For \( H \in \mathcal{H} \), we denote the corresponding realizations for \( \Omega_n \) and \( \Psi \) by \( \Omega \) and \( \Psi \), respectively. Following similar lines of reasoning as in (81), we get the inequality in (100), shown at the bottom of the next page. We note that

\[
\mathbb{E}[\mathbf{v}^T \Psi \mathbf{v}] = \text{tr}(\Psi)
\]

(a) \( \text{tr}(\Omega_n^{-1} + \Sigma_n^{1/2} H^T \Omega_n^{-1} H \Sigma_n^{1/2}) \)

(b) \( \text{tr}(\Omega_n^{-1} + \Omega_n^{-1} H \Sigma_n H^T) \)

\[
= \text{tr}(\Omega_n^{-1} (I_m + H \Sigma_n H^T))
\]

\[
= \text{tr}(\Omega_n^{-1} \Omega)
\]

\[
= \text{tr}(I_m) = m,
\]

(101)

where (a) follows by the expression for \( \Psi \) in (97) and (b) uses the identity \( \text{tr}(M_1 M_2) = \text{tr}(M_2 M_1) \) to write \( \text{tr}(\Sigma_n^{1/2} H^T \Omega_n^{-1} H \Sigma_n^{1/2}) = \text{tr}(\Omega_n^{-1} H \Sigma_n H^T) \). Then (100) becomes

\[
\Pr \left( \left| \mathbf{v}^T \Psi \mathbf{v} - m \right| \geq m \eta \right| H = H \)
\]

\[
\leq \Pr \left( \left| \mathbf{v}^T \Psi \mathbf{v} - \mathbb{E}[\mathbf{v}^T \Psi \mathbf{v}] \right| \geq m \eta \right)
\]

\[
\leq \Pr \left( \frac{\text{Var}(\mathbf{v}^T \Psi \mathbf{v})}{m^2 \eta^2} \right)
\]

\[
= \frac{2 \text{tr}(\Psi^2)}{m^2 \eta^2},
\]

(102)

where (a) is due to Chebyshev’s inequality and (b) is due to \( \text{Var}(\mathbf{v}^T \Psi \mathbf{v}) = 2 \text{tr}(\Psi^2) \) which relies on the fact that \( \mathbf{v} \) is Gaussian. It is shown in Appendix F that

\[
\text{tr}(\Psi^2) \leq C_n'/2,
\]

(103)

where \( C_n' \) is defined in (93). Then

\[
\Pr \left( \left| \mathbf{v}^T \Psi \mathbf{v} - m \right| \geq m \eta \right| H = H \leq \frac{C_n'}{m^2 \eta^2}.
\]

(104)

The right side of (104) does not depend on the choice of \( H \in \mathcal{H} \). Then (98), (99) and (104) complete the proof. ■

By the previous proposition, if \( \lim_{n \to \infty} \frac{1}{n} \lambda_{\max}(\Sigma_n) = 0 \) holds, then

\[
\lim_{n \to \infty} \Pr \left( \mathbf{y} \notin T^m(\nu)(\Omega_n) \right) = 0,
\]

(105)
for every $\eta' > 0$.

Next, we concentrate on the second term on the right side of (91). Conditioning on $H = h$ and denoting the corresponding realization of $\Omega_n$ by $\Omega_n$, we obtain the upper bound in (106), shown at the bottom of the next page, where (a) is due to independence of $x_i$ and $y$ for $i \neq 1$ and (b) is due to

$$
p_{x_i}(x) = p_G(x; \Sigma_n) \leq 2^{-h_G(\Sigma_n)\log n} \quad (107)$$

for $x \in T^{(n)}_\varepsilon(\Sigma_n)$ and

$$
p_{y_i|H}(y|H) = p_G(y_i; \Omega_n) \leq 2^{-h_G(\Omega_n)\log n} \quad (108)$$

for $y \in T^{(m)}_\eta(\Omega_n)$. To proceed, we need to find an upper bound on the third term on the right side of (106), i.e., the volume of a Gaussian typical set. Reference [10] provides the upper bound $2^{2h_G(\Sigma_n)\log n^2\eta}$ on $T^{(m+n)}_\eta(\Sigma_n)$. This upper bound turns out to be quite loose for our purposes. The parameter $\eta$ will eventually be replaced by $\eta_n$ in (70) which scales with $P$. A tight upper bound is provided by the next lemma.

**Lemma 1:** Let $\Sigma$ be an $n \times n$ positive-definite matrix and $\eta > 0$. Then

$$\text{Vol}(T^{(n)}_\eta(\Sigma)) \leq 2^{h_G(\Sigma)\log(1+\eta)} \quad (109)$$

This upper bound is tight in the sense that for every sequence of $n \times n$ positive-definite matrices $\Sigma_n$,

$$\lim_{n \to \infty} \left( \frac{\text{Vol}(T^{(n)}_\eta(\Sigma_n))}{2^{h_G(\Sigma)\log(1+\eta)}} \right) = 1 + \eta. \quad (110)$$

**Proof:** See Appendix I.

Applying (109) to the $(m+n) \times (m+n)$ matrix $\Xi_n$,

$$\text{Vol}(T^{(m+n)}_\eta(\Xi_n)) \leq 2^{h_G(\Xi_n)\log(1+\eta)} \quad (111)$$

By (106) and (111), we get the inequality in (112), shown at the bottom of the next page. By (67),

$$h_G(\Xi_n) - h_G(\Sigma_n) = \frac{1}{2} \log((2\pi e)^{m+n} \det(\Xi_n)) - \frac{1}{2} \log((2\pi e)^n \det(\Sigma_n)) = \frac{m}{2} \log(2\pi e). \quad (113)$$

Using this and writing $h_G(\Omega_n) = \frac{m}{2} \log(2\pi e) + \frac{1}{2} \log \det(\Omega_n)$, we see that the expression on the right side of (112) is equal to

$$2^{-\frac{1}{2} \log \det(\Omega_n) + \frac{m}{2} \log(2\pi e) + \frac{1}{2} \log(\det(\Omega_n))} \quad (114)$$

Averaging over $H \in \mathcal{H}$ according to the probability law for the random matrix $\mathcal{H}$, we arrive at the inequality

$$\Pr \left( \mathcal{E}_1 \in T^{(m)}_\varepsilon(\Sigma_n), \mathcal{E}_2 \in T^{(m+n)}_\eta(\Xi_n), y \in T^{(m)}_\eta(\Omega_n) \right) \leq \mathbb{E} \left[ 2^{-\frac{1}{2} \log \det(\Omega_n)} \right] \times 2^{\frac{m}{2} \log(1+\eta) + \frac{1}{2} \log(2\pi e) + \frac{1}{2} \log(\det(\Omega_n))}. \quad (115)$$

Since the distribution of $H$ is unknown, the expectation on the right side of (115) can not be computed. The next lemma provides an almost-sure lower bound on $\det(\Omega_n)$.

**Lemma 2:** Recall $\gamma_n$ in (23). Then

$$\det(\Omega_n) \geq (1 - \gamma_n)^n \det(I_m + H \Sigma_n H^T) \quad (116)$$

holds almost surely.

**Proof:** The proof relies on Weyl's inequality on perturbation of eigenvalues of a symmetric matrix. See Appendix J for the details.

Using Lemma 2, we get the upper bound

$$\mathbb{E} \left[ 2^{-\frac{1}{2} \log \det(\Omega_n)} \right] \leq 2^{-\frac{1}{2} \log(1+\gamma_n) + \frac{1}{2} \log(2\pi e) + \frac{1}{2} \log(\det(\Omega_n))}. \quad (117)$$

Define

$$\tilde{\Omega}_n := I_m + H \Sigma_n H^T \quad (118)$$

By (115), (117) and (118), we reach (119), shown at the bottom of the next page, where $\delta_n$ is given by

$$\tilde{\delta}_n := \frac{m}{2n} \log(1+\gamma_n) + \frac{1}{2n} \log(2\pi e) + \frac{m}{2n} \log(\det(\Omega_n)). \quad (120)$$

This is subject to the conditions $\eta > \eta_n$, $\eta' > 0$. Letting $\eta$, $\eta'$ and $\varepsilon$ approach $\eta_n$, 0 and 0 from above, respectively, we see that the right side of (119) vanishes as $n$ grows large if

$$R \leq \lim_{n \to \infty} \left( \frac{1}{2n} \log \det(\tilde{\Omega}_n) - \log(1+\gamma_n) - \delta_n \right), \quad (121)$$

where $\delta_n$ is given in (25). The proof of Theorem 1 is now complete.

### VI. Achievable Rates When the Receiver Knows the Realization of $H$

In this section, we assume the receiver knows the realization $H$ for the random channel matrix $H$. We denote the capacity in this scenario by $C_{\text{CSIR}}^C$. A possible decoding scheme is the one in (64) with $\Xi_n$ replaced by $\tilde{\Xi}_n$ given by

$$\tilde{\Xi}_n := \left[ \frac{\Sigma_n}{H^T \Sigma_n I_m + H \Sigma_n H^T} \right]. \quad (122)$$

i.e., $\tilde{\Xi}_n$ is the matrix $\Xi_n$ after one replaces $H$ by $H$. In this case, the Gaussian joint typicality decoder is tuned to the actual realization of the channel matrix $H$.

Error analysis is very similar to the error analysis presented in Section V. The probability of the type I error is bounded by

$$\Pr(\text{Error}^{(I)}) \leq \Pr \left( \mathcal{E}_1 \notin T^{(m-n)}_\varepsilon(\Sigma_n) \right) + \Pr \left( \mathcal{E}_2 \notin T^{(m+n)}_\eta(\Xi_n) \right). \quad (123)$$

The first term on the right side of (123) is exactly the first term on the right side of (69) and vanishes as $n$ increases regardless of $\varepsilon > 0$. To find an upper bound on the second term in (123), assume the following

$$\Pr \left( \| v^T \Xi v - m \| \geq m' \right) \leq \Pr \left( \| v^T \Xi v - \mathbb{E}[v^T \Xi v] \| \geq m'(\eta' - |\mathbb{E}[v^T \Xi v]/m - 1|) \right). \quad (100)$$

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
term $\Pr(w_i \notin T^{(m+n)}(\Xi_n))$, we follow similar lines of reasoning in Proposition 2. Since the decoder is tuned to the actual channel matrix, the error matrix $E$ in (78) is the zero matrix and hence, the matrix $\Phi$ in (77) turns into the identity matrix $I_{m+n}$. Then (79) is replaced by
\[
\Pr \left( \mathbf{w}_i \notin T^{(m+n)}(\Xi_n) \right) = \Pr \left( \frac{1}{m+n} \mathbf{v}^T \nu - 1 \geq \eta \right).
\] (124)

Note that $\mathbf{v}^T \nu = \|\mathbf{v}\|^2$ is a $\chi^2_{m+n}$ random variable. The right side of (124) vanishes as $n$ increases due to the weak law of large numbers. This holds regardless of the value of $\eta > 0$. We see that $\Pr(\text{Error}(I))$ vanishes regardless of $\epsilon, \eta > 0$ and regardless of the choice of the covariance matrix $\Sigma_n$. This is in contrast to the result of Proposition 2 which shows that in the absence of $H$ at the receiver end, $\Pr(\text{Error}(I))$ vanishes if $\eta$ is sufficiently large as given in (70) and $\Sigma_n$ satisfies the condition in (21).

As for the type II error, the lines of reasoning are exactly those presented in Section V for the scenario where $H$ is unknown at the receiver except for the fact that $\eta > 0$ is now arbitrary. We still require $\Sigma_n$ to satisfy the condition in (21) so that the right side of (92) is guaranteed to vanish as $n$ grows large. Letting $\eta, \eta', \epsilon$ all approach zero from above, the probability of the type II error vanishes exponentially in $n$ if
\[
R < \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{E} \left[ 2^{-\frac{1}{2} \log \det(\Omega_n)} \right].
\] (125)

Computing this limit is challenging even in a scenario where the channel taps vary independently. Using the result of Lemma 2, one can loosen the bound in (125) as
\[
R < C^*_{\text{CSI}}, \quad \text{LB} := \liminf_{n \to \infty} \left( \frac{1}{2n} \log \det(I_n + H^T H \Sigma_n) - \delta_n \right),
\] (126)

\(^8\)A chi-squared random variable with $m+n$ degrees of freedom.

where $\delta_n$ is given in (25). Maximizing $C^*_{\text{CSI}}, \text{LB}$ over $\Sigma_n$ subject to the conditions in (20) and (21), we obtain
\[
\sup_{\Sigma_n \text{ subject to (21) and (22)}} C^*_{\text{CSI}}, \text{LB}
= \frac{1}{4\pi} \int_0^{2\pi} \log(\Theta_1(\omega)) \, d\omega - \phi_1,
\] (127)

where $\Theta_1$ satisfies (45) and $\phi_1$ is given in (49). The first term on the right side of (127) is the capacity $C^*_0$ in the absence of channel variations. It follows that
\[
C^*_0 - C^*_{\text{CSI}} \leq \phi_1.
\] (128)

If $\alpha > 0$, then $\phi_1$ is bounded in the transmission power $P$. In this case, we see that in the presence of channel state information at the receiver, one can achieve $C^*_0$ within a constant that does not depend on $P$.

VII. SUMMARY AND CONCLUDING REMARKS

We studied a stochastic and time varying Gaussian ISI channel where the $i$th tap during time slot $t$ is a random variable whose probability law is supported over an interval of radius $r_i$ centred at $c_i$. The joint distribution as well as realizations of the array of channel taps are unknown to both ends of communication. A lower bound was derived on the channel capacity by carefully working out the details of error analysis for a decoder which functions based on Gaussian joint-typicality decoding tuned to the matrix $H_c$. The proposed lower bound saturates at a positive level as $P$ increases beyond the saturation power $P_{\text{sat}}$ given in (58). In a scenario where perfect channel state information is available at the receiver, the capacity lies within a gap of size $\phi_1$ given in (49) to the capacity in the absence of channel variations. If the central channel frequency response is everywhere nonzero, then $\alpha > 0$ and the gap $\phi_1$ does not scale with $P$.

A partial converse result was presented in a scenario where different channel taps evolve independently of each other and each tap process is stationary with a finite differential entropy rate. It shows that for a sequence of codebooks with vanishingly small probability of error, if $\tau = \frac{1}{2} \log P + O(1)$...
where $\tau$ is defined in (34), then the rate of that sequence of codebooks does not scale with $P$.

An observation made in the paper which may find other applications is a tight upper bound on the volume of Gaussian typical sets given in Lemma 1.

We close this section by mentioning that the decoding rule adopted in this paper is reminiscent of the notion of “mismatched decoding” which is extensively studied for memoryless channels [20]. Characterizing fundamental limits of mismatched decoding for time-varying channels with memory is another direction for investigation.

APPENDIX A

$C^s$ IS UNBOUNDED IN P IF $r_i < |c_i|/2$ FOR SOME $i$

To avoid ISI, one transmits a sequence of $k$ zeros after sending each code symbol. To keep the presentation simple, let $k = 1$. Then every other transmitted signal is a zero. Let the symbols for the transmitted codeword be $x_1, \ldots, x_n$ and denote the transmitted signal and the received signal during time slot $t$ by $x_t$ and $y_t$, respectively. We have

$$x_t = \begin{cases} 
\hat{x}_i & t = 2i - 1, \quad i = 1, 2, \ldots, n \\
0 & t = 2i,
\end{cases}$$

and

$$y_t = \begin{cases} 
h_{t,0} \hat{x}_i + z_t & t = 2i - 1, \quad i = 1, 2, \ldots, n, \\
h_{t,1} \hat{x}_i + z_t & t = 2i.
\end{cases}$$

Fix an integer $M > 0$. Building on an idea put forth in Section VI-B in [8], we show that a number of $\frac{\log M}{L}$ bits per time slot can be transmitted reliably as $P$ grows to infinity. Consider a transmitter that uses a real constellation $\mathcal{A} = \{a_1, \ldots, a_M\}$ where $a_1 < \cdots < a_M$. For an uncoded transmitter, we have $n = 1$ and the only transmitted signal is $\hat{x}_1 \in \mathcal{A}$ which we denote by $x$ for notational simplicity. The received signals are

$$y_1 = h_{1,0} x + z_1, \quad y_2 = h_{2,1} x + z_2,$$

where $h_{1,0}$ and $h_{2,1}$ are random variables with supports $[c_0 - r_0, c_0 + r_0]$ and $[c_1 - r_1, c_1 + r_1]$, respectively. Consider a suboptimal nearest neighbour decoder$^{10}$ that performs only one of the following alternatives:

1) It decides $a_i \in \mathcal{A}$ is transmitted if $c_0 a_i$ is the closest number to $y_1$. The signal $y_1$ is discarded.

2) It decides $a_i \in \mathcal{A}$ is transmitted if $c_1 a_i$ is the closest number to $y_2$. The signal $y_2$ is discarded.

Let us consider the first alternative. We shall denote $c_0 r_0, h_{1,0}, z_1$ and $y_1$ by $c, r, h, z$ and $y$, respectively, for notational simplicity. The estimate $\hat{x}$ for $x$ produced by this decoder is given in (132), shown at the bottom of the next page. The probability of decoding error is written in (133), shown at the bottom of the next page. Choose $a_i = L^i$ for $i = 1, \cdots, M$ where$^{11}$ $L > 0$ is a constant that satisfies the power constraint $\frac{1}{M} \sum_{i=1}^{M} L^{2i} = P$. This gives $\frac{1}{M} L^{2(M-1)2} = P$ and hence, we need $L \approx (MP)\frac{1}{2}$ for large $P$. We show that if $c > 0$ and $r < \frac{c}{2}$, then each of the conditional probabilities on the right side of (133) tends to one as $L$ grows large. Fix $i = 2, \cdots, M - 1$. Then

$$\Pr \left( \frac{c a_{i-1} + c a_i}{2} < y < \frac{c a_i + c a_{i+1}}{2} \mid x = a_i \right) \leq \Pr \left( \frac{c a_{i-1} + c a_i}{2} < h a_i + z < \frac{c a_i + c a_{i+1}}{2} \right) \leq \Pr \left( \frac{c(a_{i-1} + a_i)}{2a_i} < h + \frac{z}{a_i} < \frac{c(a_i + a_{i+1})}{2a_i} \right) = \Pr \left( \frac{c(1/L + 1)}{2} < h + z/L^i < \frac{c(L+1)}{2} \right).$$

We write

$$\Pr(\mathcal{E}_L) = \Pr(h + \frac{z}{L^i} > \frac{c}{2L}).$$

The sequence of random variables$^{12}$ $h + \frac{z}{L^i} - \frac{c}{2L}$ converges almost surely to the random variable $h$ as $L \to \infty$. Hence, $h + \frac{z}{L^i} - \frac{c}{2L}$ also converges weakly$^{13}$ to $h$. The topological boundary of the interval $(c/2, \infty)$ is $\{c/2\}$. Since $r < c/2$, we have $h + \frac{z}{L^i} = 0$ and (136) gives

$$\lim_{L \to \infty} \Pr(\mathcal{E}_L) = \Pr(h > c/2) = 1.$$ 

Similarly, we write

$$\Pr(\mathcal{F}_L) = \Pr\left( \frac{h}{L} + \frac{z}{L+1} - \frac{c}{2L} < \frac{c}{2} \right).$$

The sequence of random variables $\frac{h}{L} + \frac{z}{L+1} - \frac{c}{2L}$ converges almost surely to 0 as $L \to \infty$. Hence, $\frac{h}{L} + \frac{z}{L+1} - \frac{c}{2L}$ also converges weakly to 0. The topological boundary of the interval $(-\infty, c/2)$ is $\{c/2\}$ and $\Pr(0 = c/2) = 0$. Therefore,

$$\lim_{L \to \infty} \Pr(\mathcal{F}_L) = \Pr(0 < c/2) = 1.$$ 

By (137) and (139),

$$\lim_{L \to \infty} \Pr(\mathcal{E}_L \cap \mathcal{F}_L) = 1,$$

due to the inequality $\Pr(\mathcal{E}_L \cap \mathcal{F}_L) \geq 1 - \Pr(\mathcal{E}_L^c) - \Pr(\mathcal{F}_L^c)$. Similarly, the last two terms on the right side of (133) tend to one as $L$ increases.

$^{9}$Extension to arbitrary $k \geq 2$ is straightforward.

$^{10}$Maximum likelihood decoding does not really apply as the distributions of $h_{1,0}$ and $h_{2,1}$ are unknown.

$^{11}$For simplicity of presentation, we have ignored the removal of the transmitted DC signal.

$^{12}$Assume $L$ takes on integers.

$^{13}$Recall a sequence of random variables $x_n$ is said to converge weakly to a random variable $x$ if $\lim_{n \to \infty} \Pr(x_n \in A) = \Pr(x \in A)$ for every Borel set $A$ such that $\Pr(x \in \text{bd}(A)) = 0$. Here, $\text{bd}(A)$ denotes the topological boundary of $A$. The fact that almost sure convergence implies weak convergence appears in most texts on probability, e.g., [14].
APPENDIX B
PROOF OF THEOREM 2

Let $C_n = \{z_{n,1}, \ldots, z_{n,2^nR}\}$ be a sequence of codebooks with rate $R$ and vanishingly small probability of error. Throughout this appendix, we will drop the codebook index $n$ and denote $z_i$ by $z_i$ for notational simplicity. The $j^{th}$ symbol of $z_i$ is denoted by $x_{i,j}$. The codebooks satisfy the average transmission power constraint in (31). The transmitted books with rate $R$ lead to a sequence of codebooks.

We write $I(x; y) = h(y) - h(y|x)$ and find an upper bound and a lower bound on the terms $h(y)$ and $h(y|x)$, respectively.

A. Upper Bound on $h(y)$

To obtain an upper bound on $h(y)$, we follow similar lines of reasoning in [9]. Let $q(\cdot)$ be a given PDF on $\mathbb{R}^m$ which is positive everywhere. Recall the relative entropy (Kullback-Leibler divergence) between $p_y(\cdot)$ and $q(\cdot)$ is given by

$$D_{KL}(p_y||q) := \int_{\mathbb{R}^m} p_y(y) \log \frac{p_y(y)}{q(y)} dy = -h(y) - \int_{\mathbb{R}^m} p_y(y) \log q(y) dy.$$  

Since $D_{KL}(p_y||q) \geq 0$, we obtain the upper bound

$$h(y) \leq -\int_{\mathbb{R}^m} p_y(y) \log q(y) dy. \quad (143)$$

This provides a myriad of bounds on $h(y)$ for different choices of $q(\cdot)$. A particularly interesting choice is given by

$$q(y) = \frac{\Gamma(n)}{\pi^m b^n} \exp \left( -\frac{\|y\|^2}{b} \right), \quad (144)$$

for some constant $b > 0$ where $\Gamma(\cdot)$ is the Gamma function. The density in (144) is a special case of a multivariate density.

Construct a vector $\bar{x}$ such that its $j^{th}$ element $x_{i,j}$ is in $(c_{ai-1} + c_{ai+1})/2$ and $i = 2, \ldots, M - 1$.

$$\bar{x} = \begin{cases} a_i & y \in \left(\frac{c_{ai-1} + c_{ai+1}}{2}, \frac{c_{ai+1} + c_{ai+2}}{2}\right), \quad i = 2, \ldots, M - 1 \\ a_1 & y < \frac{c_{a_1} + c_{a_2}}{2} \\ a_M & y > \frac{c_{a_{M-1}} + c_{a_M}}{2}. \end{cases} \quad (132)$$

$$\Pr(\bar{x} \neq x) = 1 - \Pr(\bar{x} = x)$$

$$= 1 - \frac{1}{M} \sum_{i=2}^{M-1} \Pr \left( \frac{c_{ai-1} + c_{ai+1}}{2} < y < \frac{c_{ai} + c_{ai+2}}{2} | x = a_i \right)$$

$$- \frac{1}{M} \Pr \left( y < \frac{c_{a_1} + c_{a_2}}{2} | x = a_1 \right)$$

$$- \frac{1}{M} \Pr \left( y > \frac{c_{a_{M-1}} + c_{a_M}}{2} | x = a_M \right). \quad (133)$$
\[ h(y) \leq \frac{m}{2} \log \pi - \log \Gamma(m/2) + (m/2 - 1) \log m \]
\[ + \frac{m/2 - 1}{2nR} \sum_{i=1}^{2nR} \log (2(\beta + r_s)^2 ||x_i||_2^2/n + 2) \]
\[ + \log \mathbb{E}[\|y\|^2_2] + \frac{1}{\ln 2} \]  
\[ \text{for every } i = 1, \ldots, 2nR, \]
\[ h(y_i) = h(Hx_i + z) = h(Hx_i + z), \]  
\[ \text{where the last step is due to independence of } x \text{ and the pair } (H, z). \]
\[ y_i := Hx_i + z. \]

The random matrix $H$ does not have a density as many of its entries are zeros. Hence, it is meaningless to talk about the differential entropy of $H$. To avoid this difficulty, let $h_i$ be a vector of length $(k + 1)n$ that results by placing the vectors $h_0, h_1, \ldots, h_k$ beneath each other, i.e.,
\[ h := (h_0^T, h_1^T, \ldots, h_k^T)^T. \]  

Then the vector $Hx_i$ can be written as
\[ Hx_i = X_i h_i, \]
where $X_i$ is an $m \times (k + 1)n$ matrix whose entries depend on the entries of the codeword $x_i$. Let us denote the entries of $x_i$ by $x_{i,1}, \ldots, x_{i,n}$. To describe the matrix $X_i$ in (159), it is convenient to define the diagonal matrix $D_i$ by
\[ D_i := \text{diag}(x_{i,1}, \ldots, x_{i,n}), \]
\[ i.e., D_i \text{ is an } n \times n \text{ diagonal matrix that carries the entries of the codeword } x_i \text{ on its main diagonal. The matrix } X_i \text{ contains } k + 1 \text{ copies of the matrix } D_i \text{ as its submatrices. To demonstrate this, consider the case where } n = 4 \text{ and } k = 2. \]

We also denote the set of indices among $1, 2, \ldots, m$ that do not belong to $A_l$ by $A^c_l$, i.e.,
\[ A^c_l = \{1, 2, \ldots, m\} \setminus A_l. \]

Recall the vector $y_i$ in (157). Let $(y_{i,j}, A_i)$ be the subvector of $y_i$ which contains those entries of $y_i$ that lie at positions inside the index set $A_i$. Similarly, we define $(y_{i,j}, A^c_l)$. The main observation is that
\[ (y_{i,j}, A_l) = (Hx_i, A_l) + (z, A_l) = (X_i h_i + \tilde{z}, A_l), \]
\[ = D_i h_i + \tilde{z}, \]
\[ \text{for every } l = 0, 1, \ldots, k \text{ where the vector } \tilde{z} \text{ only depends on } h_j \text{ for } j \neq l. \]

To demonstrate this, once again consider the case where $n = 4$ and $k = 2$. Here, $l$ takes on $0, 1, 2$. The matrix $X_i$ is given in (161). We have $A_0 = \{1, 2, 3, 4\}$. Then
\[ (X_i h), A_0 = D_i h_0 + \tilde{z}, \]  
\[ \text{where } \tilde{z} \text{ is the vector defined by } \tilde{z} \text{ in (157)} \text{ and } A_0 \text{ in (163)}. \]

The vectors $h_i$ are defined in (26).
Similarly, one can find \( \theta_i \) and \( \vartheta_j \) explicitly. Now, let us fix \( l_0 = 1, \cdots, k \). By the chain rule for differential entropy [17],

\[
h(y_i) = h((y_i)_{A_l}) + h((y_i)_{A_l^c} | (y_i)_{A_l}).
\]

We write

\[
\begin{align*}
(h((y_i)_{A_l}) &= h(D_i h_i + \vartheta_i + (z)_{A_l}), \\
&\geq h(D_i h_i + \vartheta_i + (z)_{A_l} | \vartheta_i), \\
&\geq h(D_i h_i + (z)_{A_l}), \\
&\geq \frac{n}{2} \log \left( 2^{\frac{d}{2}} h(D_i h_i) + 2^{\frac{d}{2}} h((z)_{A_l}) \right), \\
&= \frac{n}{2} \log \left( 2^{\frac{d}{2}} \log |\det(D_i)| + h(D_i h_i) + 2\pi e \right), \\
&= \frac{n}{2} \log (2\pi e) \\
&+ \frac{n}{2} \log \left( 1 + \frac{2^{\frac{d}{2}} h(D_i h_i)}{2\pi e} |\det(D_i)| \right),
\end{align*}
\]

(167)

where \((a)\) is due to (164), \((b)\) is due to the fact that conditioning does not increase the differential entropy, \((c)\) is due to the vectors \(D_i h_i + (z)_{A_l}\) and \(\vartheta_i\) being independent, \((d)\) is due to the vertex entropy power inequality [17] and the fact that both vectors \(D_i h_i\) and \((z)_{A_l}\) have densities, \((e)\) is due to \(h((z)_{A_l}) = \frac{d}{2} \log (2\pi e)\) and the scaling property for differential entropy which states \(h(A \xi) = \log |\det(A)| + h(\xi)\) for an invertible matrix \(A\) and a random vector \(\xi\) and in \((f)\) we have factored \(2\pi e\) from both terms inside the logarithm. Moreover,

\[
\begin{align*}
(h((y_i)_{A_l^c} | (y_i)_{A_l}) &= h((H x_i)_{A_l^c} | (y_i)_{A_l}) \\
&\geq h((H x_i)_{A_l^c} | (y_i)_{A_l}, (H x_i)_{A_l}) \\
&\geq h((z)_{A_l}), \\
&\geq \frac{m-n}{2} \log (2\pi e),
\end{align*}
\]

(168)

where \((a)\) is due to the fact that conditioning does not increase the differential entropy, \((b)\) holds due to independence of \((z)_{A_l^c}\) and the pair \(((y_i)_{A_l}, (H x_i)_{A_l})\) and \((c)\) is due to \((z)_{A_l^c} \sim N(0, I_{m-n})\). Putting (166), (167) and (168) together,

we get

\[
h(y_i) \geq \frac{m}{2} \log (2\pi e) \\
+ \frac{n}{2} \log \left( 1 + \frac{2^{\frac{d}{2}} h(D_i h_i)}{2\pi e} |\det(D_i)| \right).
\]

This is true for every \(l = 1, \cdots, k\). Hence,

\[
h(y_i) \geq \frac{m}{2} \log (2\pi e) \\
+ \frac{n}{2} \log \left( 1 + \frac{2^{\frac{d}{2}} h(D_i h_i)}{2\pi e} |\det(D_i)| \right),
\]

(170)

where \(\rho_n\) is defined in (30). By (155), (156), (157) and (170),

\[
h(y_i) \geq \frac{m}{2} \log (2\pi e) \\
+ \frac{n}{2} \sum_{i=1}^{2n} \log \left( 1 + \frac{2^{\frac{d}{2}} h(D_i h_i)}{2\pi e} |\det(D_i)| \right).
\]

By (154) and (171), the proof of (32) is complete.

\section*{Appendix C}
\section*{Proof of Proposition I}

Fix \(\varepsilon > 0\). We maximize \(\frac{1}{2m} \sum_{i=1}^{n} \log (1 + \lambda_{n,i} d_{n,i}) - \log (1 + (k+1) \|\xi\|^2 - \sum_{i=1}^{n} d_{n,i})\) subject to the conditions \(d_{n,i} \geq \varepsilon\) and \(\sum_{i=1}^{n} d_{n,i} \leq n \bar{P}\). The Lagrangian is given by

\[
\mathcal{L} = \frac{1}{2n} \sum_{i=1}^{n} \log (1 + \lambda_{n,i} d_{n,i}) - \log (1 + (k+1) \|\xi\|^2 - \sum_{i=1}^{n} d_{n,i}) + \mu \widetilde{P} - \sum_{i=1}^{n} d_{n,i},
\]

(172)

where \(\mu_1, \cdots, \mu_n, \mu \geq 0\) are the Lagrange multipliers and the complementary slackness conditions

\[
\forall i, \quad \mu_i (d_{n,i} - \varepsilon) = 0, \quad \mu (n \bar{P} - \sum_{i=1}^{n} d_{n,i}) = 0
\]

(173)

hold. The first order necessary conditions \(\frac{\partial \mathcal{L}}{\partial d_{n,j}} = 0\) for \(j = 1, \cdots, n\) give

\[
\frac{1}{\lambda_{n,j} - d_{n,j}} = \frac{1}{1 + (k+1) \|\xi\|^2 - \sum_{i=1}^{n} d_{n,i}} + \mu_j.
\]

(174)

For all such \(j\) such that \(d_{n,j} > \varepsilon\), we must have \(\mu_j = 0\) and then (174) shows that \(\frac{1}{\lambda_{n,j}} + d_{n,j}\) does not depend on the index \(j\), i.e., it is a constant \(\Theta\) and we get

\[
d_{n,j} = \max \left( \Theta - \frac{1}{\lambda_{n,j}}, \varepsilon \right), \quad j = 1, \cdots, n.
\]

(175)
Two situations can happen as we describe next:

1. Assume $\sum_{i=1}^{n} d_{n,i} = nP$. Then $\Theta$ solves the equation

$$\sum_{i=1}^{n} \max\left(\Theta - \frac{1}{\lambda_{n,i}}, \varepsilon\right) = nP.$$ 

Dividing both sides by $n$ and letting $n$ grow to infinity, Szegö’s Theorem gives

$$\frac{1}{2\pi} \int_{0}^{2\pi} \max\left(\Theta - |f(\omega)|^{-2}, \varepsilon\right)d\omega = P. \quad (176)$$

2. Assume $\sum_{i=1}^{n} d_{n,i} < nP$. Then $\mu = 0$ and (174) gives

$$\frac{1}{\lambda_{n,j}} + d_{n,j} - \frac{1}{2\ln n} + \frac{k+1}{n(i+n+1)\|z_i\|_2^2} + \mu_j = 0. \quad (177)$$

If there is at least one $j$ such that $d_{n,j} > \varepsilon$, then $\mu_j = 0$ and (177) further simplifies to

$$\sum_{i=1}^{n} \left(\Theta - \frac{1}{\lambda_{n,i}}, \varepsilon\right) = 2n\Theta - \frac{m+n}{(k+1)\|z_i\|_2^2}. \quad (178)$$

Dividing both sides by $n$, letting $n$ grow to infinity and invoking Szegö’s Theorem,

$$\frac{1}{2\pi} \int_{0}^{2\pi} \max\left(\Theta - |f(\omega)|^{-2}, \varepsilon\right)d\omega = 2\Theta - \frac{2}{(k+1)\|z_i\|_2^2}. \quad (179)$$

We also need to determine $\lim_{n \to \infty} \delta_n$. This requires computing the limiting values for $d_{n,1} = \max(\Theta - \frac{1}{\lambda_{n,1}}, \varepsilon)$ and $d_{n,n} = \max(\Theta - \frac{1}{\lambda_{n,n}}, \varepsilon)$. We have $\lambda_{n,1} = \lambda_{n,n} = \lambda_{\max}(H_c^T H_c)$ and $\lambda_{n,n} = \lambda_{\max}(H_c^T H_c)$. The spectral function for the Toeplitz matrix $H_c^T H_c$ is $|f(\omega)|^2$. Then Corollary 4.2 on page 58 in [16] gives

$$\lim_{n \to \infty} \lambda_{n,1} = \alpha^2, \quad \lim_{n \to \infty} \lambda_{n,n} = \beta^2. \quad (180)$$

It is also clear that the condition in (41) holds. In fact,

$$\lim_{n \to \infty} \frac{1}{n} \lambda_{\max}(\Sigma_n) = \lim_{n \to \infty} \frac{d_{n,n}}{n} = \lim_{n \to \infty} \frac{1}{n} \max(\Theta - \beta^{-2}, \varepsilon) = 0. \quad (181)$$

Finally, letting $\varepsilon$ approach zero from above, the proof of Proposition 1 is complete.

**APPENDIX D**

**PROOF OF (87)**

We need the following lemma:

**Lemma 3:** Let $M_1$ and $M_2$ be matrices of sizes $m \times n$ and $n \times m$, respectively. Then

$$\|M_1 M_2\|_2 \leq \|M_1\| \|M_2\|_2 \quad (182)$$

and

$$\|M_1 M_2\|_2 \leq \|M_2\| \|M_1\|_2 \quad (183)$$

**Proof:** This is Problem 5.6.20 in [12]. A proof is provided in Appendix E for completeness.

For a square block matrix $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, we have $\text{tr}(M^2) = \text{tr}(A^2 + BC + CB + D^2)$. Applying this to $\Phi$ in (77), we get (184), shown at the bottom of the next page. Using the identity $\text{tr}(M_1 M_2) = \text{tr}(M_2 M_1)$,

$$\text{tr}(\Sigma_{1/2}^n E^T \Sigma_{1/2}^n) = \text{tr}(E^T E \Sigma_n)$$

and

$$\text{tr}(E \Sigma_n E^T) = \text{tr}(E^T E \Sigma_n). \quad (185)$$

Then

$$\text{tr}(\Phi^2) = m + n + 4\text{tr}(E^T E \Sigma_n)$$

and

$$\text{tr}(E^T E \Sigma_n^2)$$

We saw in (85) that $\text{tr}(E^T E \Sigma_n) \leq (k+1)\|z_i\|_2^2 \text{tr}(\Sigma_n) \leq (k+1)nP\|z_i\|_2^2$. The term $\text{tr}(\Sigma_{1/2}^n E^T \Sigma_{1/2}^n)^2$ can be bounded as

$$\text{tr}(\Sigma_{1/2}^n E^T \Sigma_{1/2}^n)^2 \leq \|\Sigma_{1/2}^n E^T \Sigma_{1/2}^n\|_2^2$$

and

$$\text{tr}(\Sigma_{1/2}^n E^T \Sigma_{1/2}^n)^2 \leq \|\Sigma_{1/2}^n E^T \Sigma_{1/2}^n\|_2^2$$

where (a) follows the definition of the norm $\| \cdot \|_2$, (b) is due to Lemma 3, (c) uses the fact that the matrix norm $\| \cdot \|$ is sub-multiplicative, (d) is due to $\|\Sigma_{1/2}^n\|_2^2 = \text{tr}(\Sigma_n)$ and (e) is due to $\text{tr}(\Sigma_n) \leq nP$. We need the next lemma in order to find an upper bound on $\|E^T E\|^2$.

**Lemma 4:** For a matrix $M$, define the maximum row-sum norm by

$$\|M\|_r := \max_{i} \sum_{j} |M_{i,j}|. \quad (188)$$

Then $\| \cdot \|_r$ is a matrix norm. In particular, it is sub-multiplicative, i.e.,

$$\|M_1 M_2\|_r \leq \|M_1\|_r \|M_2\|_r, \quad (189)$$

for every two matrices $M_1$ and $M_2$ of proper sizes. Moreover,

$$\|M\|^2 \leq \|M^T M\|_r, \quad (190)$$

for every matrix $M$.

**Proof:** See [12]. The inequality in (190) follows from Theorem 5.6.9 in [12] which states that the size of every eigenvalue of a square matrix is bounded from above by every matrix norm of that matrix.

We write

$$\|E^T E\|^2 \leq \|E^T E\|^2_r \leq \|E\|^2 \|E^T\|^2_r, \quad (191)$$

where (a) uses (190) and (b) is due to (189). The matrix $E$ is banded, i.e., $E_{i,j} = 0$ for $i - j < 0$ or $i - j > k$ and $|E_{i,j}| \leq r_{i-j}$ for $0 \leq i - j \leq k$. It follows that

$$\|E\|_r \leq \sum_{i=0}^{k} r_i = r_s, \quad \|E^T\|_r \leq \sum_{i=0}^{k} r_i = r_s. \quad (192)$$

By (191) and (192),

$$\|E^T E\|^2 \leq r_s^4. \quad (193)$$
Then (187) and (193) give
\[\text{tr}((\Sigma_n^{1/2} E^T E\Sigma_n^{1/2})^2) \leq nP r_s^4 \lambda_{\text{max}}(\Sigma_n). \tag{194}\]

We have shown that
\[\text{tr}(\Phi^2) \leq m + n + 4(k + 1)nP \|E\|^2 + nP r_s^4 \lambda_{\text{max}}(\Sigma_n). \tag{195}\]

**APPENDIX E**

**PROOF OF LEMMA 3**

First, we verify (182). We have
\[\|M_1M_2\|^2 = \text{tr}(M_1M_2(M_1M_2)^T). \tag{196}\]

Denote the columns of \(M_2\) by \(v_1, \ldots, v_m\). Then the columns of \(M_1M_2\) are \(M_1v_1, \ldots, M_2v_m\) and we get
\[M_1M_2(M_1M_2)^T = \sum_{i=1}^m M_1v_i(M_1v_i)^T. \tag{197}\]

By (196) and (197),
\[\|M_1M_2\|^2 = \text{tr}\left(\sum_{i=1}^m M_1v_i(M_1v_i)^T\right) = \sum_{i=1}^m \text{tr}(M_1v_i(M_1v_i)^T) = \sum_{i=1}^m \|M_1v_i\|^2 \leq \sum_{i=1}^m \|M_1\|^2 \|v_i\|^2 = \|M_1\|^2 \sum_{i=1}^m \|v_i\|^2 = \|M_1\|^2 \|M_2\|^2, \tag{198}\]

where (a) is due to the definition of the operator norm and (b) is due to \(\|M_2\|^2 = \sum_{i=1}^m \|v_i\|^2\). This proves (182). Using the facts that \(\|M\| = \|M^T\|\) and \(\|M\|_2 = \|M^T\|_2\) for every matrix \(M\), we get
\[\|M_1M_2\|_2 = \|(M_1M_2)^T\|_2 = \|M_2^T M_1^T\|_2 \leq \|M_2^T\| \|M_1^T\|_2 = \|M_2\| \|M_1\|_2, \tag{199}\]

where the inequality is due to (182) that we just verified. This completes the proof of (183).

**APPENDIX F**

**PROOF OF (103)**

By (97),
\[\text{tr}(\Psi^2) = \text{tr} ((\Sigma_n^{1/2} H^T \Omega_n^{-1} H\Sigma_n^{1/2})^2) + 2\text{tr}(\Sigma_n^{1/2} H^T \Omega_n^{-1} H\Sigma_n^{1/2}) + \text{tr}(\Omega_n^{-2}). \tag{200}\]

Since each eigenvalue of \(\Omega_n^{-1}\) is at most 1, we have
\[\text{tr}(\Omega_n^{-2}) \leq m. \tag{201}\]

Moreover,
\[\text{tr}((\Sigma_n^{1/2} H^T \Omega_n^{-1} H\Sigma_n^{1/2})^2) = \|\Sigma_n^{1/2} H^T \Omega_n^{-1} H\Sigma_n^{1/2}\|^2 \leq \|\Sigma_n^{1/2} H^T \Omega_n^{-1} H\Sigma_n^{1/2}\|^2 \tag{a}\]
\[\leq \|\Sigma_n^{1/2} H^T \Omega_n^{-1} H\Sigma_n^{1/2}\|^2 \tag{b}\]
\[\leq \|\Sigma_n^{1/2}\|_2^2 \|H^T\|^2 \|\Omega_n^{-1}\|_2 \|H\| \|\Sigma_n^{1/2}\|_2 \tag{c}\]
\[\leq \|H\| \|\Omega_n^{-1}\|_2 \lambda_{\text{max}}(\Sigma_n) \text{tr}(\Sigma_n) \tag{d}\]
\[\leq nP (\beta + r_s)^4 \lambda_{\text{max}}(\Sigma_n), \tag{202}\]

where (a) is due to Lemma 3, (b) is due to \(\|\cdot\|\) being sub-multiplicative, (c) is due to \(\|H\| = \|H^T\|\) and \(\lambda_{\text{max}}(\Sigma_n) = \|\Sigma_n^{1/2}\|_2^2\) and \(\text{tr}(\Sigma_n) = \|\Sigma_n^{1/2}\|_2^2\) and (d) is due to \(\text{tr}(\Sigma_n) \leq nP, \|H\| \leq \|H_r\| + \|E\|\) and the three inequalities
\[\|\Omega_n^{-1}\|_2 \leq 1, \|H_r\| \leq \beta, \|E\| \leq r_s, \tag{203}\]

where \(r_s\) and \(\beta\) are defined in (8) and (22), respectively. The first inequality \(\|\Omega_n^{-1}\|_2 \leq 1\) is due to \(\|\Omega_n^{-1}\|_2 = \lambda_{\text{max}}(\Omega_n^{-2})\) and the fact that none of the eigenvalues of \(\Omega_n^{-2} = (I_m + H\Sigma_n H^T)^{-2}\) are larger than 1, the second inequality \(\|H_r\| \leq \beta\) is proved in Appendix G and the third inequality \(\|E\| \leq r_s\) is proved in Appendix H. Following similar lines of reasoning as in (202),
\[\text{tr}(\Sigma_n^{1/2} H^T \Omega_n^{-2} H\Sigma_n^{1/2}) = \|\Sigma_n^{1/2} H^T \Omega_n^{-1} \Sigma_n^{1/2}\|_2 \leq \|H^T\|^2 \|\Omega_n^{-1}\|_2 \|\Sigma_n^{1/2}\|_2 \leq \|H\|^2 \text{tr}(\Sigma_n) \leq (\beta + r_s)^2 nP. \tag{204}\]

By (200), (201), (202) and (204),
\[\text{tr}(\Psi^2) \leq m + 2(\beta + r_s)^2 nP + nP (\beta + r_s)^4 \lambda_{\text{max}}(\Sigma_n). \tag{205}\]

**APPENDIX G**

**PROOF OF \(\|H_r\| \leq \beta\) IN (203)**

We need the following lemma which is a restatement of Lemma 4.1 in [16]:

**Lemma 5:** Let \(T\) be an \(n \times n\) symmetric Toeplitz matrix with real entries given by \(T_{i,j} = t_{i-j}\) where \(t_{-l} = t_l\) for every \(0 \leq l \leq n-1\). Moreover, define the function \(f_T(\cdot)\) by
\[f_T(\omega) = \sum_{i=-(n-1)}^{n-1} t_i e^{\sqrt{-1} \pi \omega i}, \quad 0 \leq \omega \leq 2\pi. \tag{206}\]

Then every eigenvalue of \(T\) lies in the interval \([\min_{\omega} f_T(\omega), \max_{\omega} f_T(\omega)]\).

Then
\[\text{tr}(\Phi^2) = \text{tr}((I_n + \Sigma_n^{1/2} E^T E\Sigma_n^{1/2})^2 + \Sigma_n^{1/2} E^T E\Sigma_n^{1/2} + E\Sigma_n E^T + I_m)\]
\[= \text{tr}(I_n + 3\Sigma_n^{1/2} E^T E\Sigma_n^{1/2} + (\Sigma_n^{1/2} E^T E\Sigma_n^{1/2})^2 + E\Sigma_n E^T + I_m). \tag{184}\]
The matrix $H_T^T H_c$ is a symmetric Toeplitz matrix with entries given by $[H_T^T H_c]_{i,j} = \sum_{i=0}^{k-|i-j|} c_i e^{\sqrt{-1} \omega i j}$. Then $f_{H_T^T H_c}(\omega) = \left| \sum_{i=0}^{k} c_i e^{\sqrt{-1} \omega i} \right|^2$. By Lemma 5, every eigenvalue $\lambda$ of $H_T^T H_c$ satisfies

$$\alpha^2 \leq \lambda \leq \beta^2,$$

(207)

where $\alpha$ and $\beta$ are given in (22). Since $\|H_c\|^2 = \lambda_{\max}(H_T^T H_c)$, then (207) gives $\|H_c\| \leq \beta$.

**Appendix H**

**Proof of $\|E\| \leq r_s$ in (203)**

The proof of $\|E\| \leq r_s$ follows similar lines of reasoning that led to $\|ET\| \leq r_s^2$ in (193). We present the details here for completeness. By Lemma 4,

$$\|E\| \leq (\|E\|_r \|ET\|_r)^{1/2},$$

(208)

where $\|\cdot\|_r$ is the maximum row-sum norm defined in (188). By (192), both $\|E\|_r$ and $\|ET\|_r$ are not larger than $r_s$. Then (208) implies that $\|E\|$ is also not larger than $r_s$.

**Appendix I**

**Proof of Lemma 1**

For $r > 0$, positive integer $n$ and real positive-definite matrix $\Sigma$, consider the ellipsoid

$$E_r := \left\{ a \in \mathbb{R}^n : \frac{1}{n} a^T \Sigma^{-1} a < r \right\}.$$  

(209)

Let $\Sigma = U \Lambda U^T$ be the spectral decomposition for $\Sigma$ where $U$ is a real orthogonal matrix and $\Lambda$ is a diagonal matrix whose diagonal entries $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\Sigma$. The set $U^T E_r := \{ U^T a : a \in E_r \}$ is an ellipsoid in standard form, i.e.,

$$U^T E_r = \left\{ a \in \mathbb{R}^n : \frac{1}{n} \sum_{i=1}^{n} \frac{a^2_i}{\lambda_i} < r \right\}.$$  

(210)

The volume of the ellipsoid $\{ (a_1, \ldots, a_n) : \sum_{i=1}^{n} \frac{a^2_i}{\lambda_i} \leq 1 \}$ is given by

$$\frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2} + 1\right)} \prod_{i=1}^{n} c_i \Gamma(\cdot)$$

(211)

where $c_i$ is the $i$th eigenvector of $\Sigma$ and $\Gamma(\cdot)$ is the Gamma function. Hence,

$$\text{Vol}(E_r) = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2} + 1\right)} \prod_{i=1}^{n} (nr \lambda_i)^{\frac{1}{2}} \lambda_i^{-\frac{n}{2}}$$

(212)

where the indicator $\mathbb{1}_{\eta > 1}$ is 1 if $\eta > 1$ and it is 0 otherwise. This proves (109). Corollary 1.2 in [11] also implies that $\Gamma(x + 1) \geq x^e e^{-x} \sqrt{2\pi(x + 1)}$ for every $x > 0$. Using this inequality in (211), we get

$$\text{Vol}(E_r) \geq \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2} + 1\right)} \prod_{i=1}^{n} \left( n r \lambda_i \right)^{\frac{1}{2}} \lambda_i^{-\frac{n}{2}} \frac{1}{x^e e^{-x} \sqrt{2\pi(x + 1)}}$$

(213)

for completeness. By Lemma 6, the proof of (207) in Appendix G, every eigenvalue of $H_T^T H_c$ is positive and hence, this matrix is invertible. In this case, the eigenvalues of $H_T^T H_c$ are given by

$$\text{Vol}(E_r) \leq \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2} + 1\right)} \prod_{i=1}^{n} \left( n r \lambda_i \right)^{\frac{1}{2}} \lambda_i^{-\frac{n}{2}} \frac{1}{x^e e^{-x} \sqrt{2\pi(x + 1)}}$$

(214)

The second step uses the identity $\text{det}(I_n + AB) = \text{det}(I_n + BA)$ for matrices $A$ and $B$ of sizes $m \times n$ and $n \times m$, respectively.
This concludes the proof of (116). We write (220),

\[ \frac{1 + \mu_i}{1 + \lambda_i} \geq 1 - \frac{\| H^T H - H_c^T H_c \| \| \Sigma_n \|}{1 + \alpha^2 \lambda_{\min}(\Sigma_n)}. \]

where the last step is due to $\| H_c \| \leq \beta$ and $\| E \| \leq r_s$ proved in Appendix G and Appendix H, respectively. By (216), (220) and (221),

\[ \det(I_m + H \Sigma_n H^T) \geq \det(I_m + H_c \Sigma_n H_c^T) \left(1 - \left(\frac{\mu_i^2 + 2r_s\beta}{1 + \alpha^2 \lambda_{\min}(\Sigma_n)}\right)^n \right). \]

This concludes the proof of (116).

ACKNOWLEDGMENT

The author is grateful to an Associate Editor, Prof. Chandra Murthy, and the anonymous reviewers whose comments significantly improved the article.

REFERENCES

[1] J. G. Proakis and M. Salehi, Digital Communications, 5th ed. New York, NY, USA: McGraw-Hill, 2007.
[2] W. Hirt and J. L. Massey, “Capacity of the discrete-time Gaussian channel with intersymbol interference,” IEEE Trans. Inf. Theory, vol. 34, no. 3, p. 38, May 1988.
[3] S. Shamai (Shitz) and R. Laroia, “The intersymbol interference channel: Lower bounds on capacity and channel precoding loss,” IEEE Trans. Inf. Theory, vol. 42, no. 5, pp. 1388–1404, Sep. 1996.
[4] Y. Carmon and S. Shamai (Shitz), “Lower bounds and approximations for the information rate of the ISI channel,” IEEE Trans. Inf. Theory, vol. 61, no. 10, pp. 5417–5431, Oct. 2015.
[5] Y. Carmon, S. Shamai (Shitz), and T. Weissman, “Comparison of the achievable rates in OFDM and single carrier modulation with I.I.D. inputs,” IEEE Trans. Inf. Theory, vol. 61, no. 4, pp. 1795–1818, Apr. 2015.
[6] H. N. Diggavi, “On achievable performance of spatial diversity fading channels,” IEEE Trans. Inf. Theory, vol. 47, no. 1, pp. 308–325, Jan. 2001.
[7] A. J. Goldsmith and M. Médard, “Capacity of time-varying channels with causal channel side information,” IEEE Trans. Inf. Theory, vol. 53, no. 3, pp. 881–899, Mar. 2007.
[8] I. C. Aboou-Faycal, M. D. Trott, and S. Shamai (Shitz), “The capacity of discrete-time memoryless Rayleigh-fading channels,” IEEE Trans. Inf. Theory, vol. 47, no. 4, pp. 1290–1301, May 2001.
[9] A. Lapidoth and S. M. Moser, “Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels,” IEEE Trans. Inf. Theory, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
[10] T. M. Cover and S. Pombra, “Gaussian feedback capacity,” IEEE Trans. Inf. Theory, vol. 35, no. 1, pp. 37–43, Jan. 1989.
[11] N. Batir, “Inequalities for the gamma function,” Arch. Math., no. 91, pp. 554–563, 2008.
[12] R. A. Horn and C. R. Johnson, Matrix Analysis, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2013.
[13] R. Bhatia, “Matrix analysis,” Graduate Texts Math., vol. 169, 1997.
[14] E. Çınlar, “Probability and stochastics,” Graduate Texts Math., vol. 261, 2011.
[15] A. C. Rencher and G. B. Schaalje, Linear Models in Statistics, 2nd ed. Hoboken, NJ, USA: Wiley, 2008.
[16] R. M. Gray, “Toeplitz and circulant matrices: A review,” Found. Trends Commun. Inf. Theory, vol. 2, no. 3, pp. 155–239, 2005.
[17] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.
[18] Y. Polyanskiy and Y. Wu, Information Theory: From Coding to Learning. Cambridge, U.K.: Cambridge Univ. Press, Nov. 2023.
[19] R. G. Gallager, Information Theory and Reliable Communications. New York, NY, USA: Wiley, 1968.
[20] J. Scarlett, A. G. I. Fàbregas, A. Somekh-Baruch, and A. Martinez, “Information-theoretic foundations of mismatched decoding,” Found. Trends Commun. Inf. Theory, vol. 17, no. 2–3, pp. 149–401, 2020.

Kamyar Moshksar was born in Shiraz, Iran. He received the B.Sc. degree in electrical engineering from Shiraz University in 2001, the M.Sc. degree from the Sharif University of Technology in 2003, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, in 2011. He moved on to receive two other master’s degrees in quantitative finance and pure mathematics from the University of Waterloo. Since 2015, he has been a Mathematics Instructor with the Columbia College, Vancouver, BC, Canada.

19This identity is part of Problem III.6.14 on page 78 in [13].